

# Population genetics on islands connected by an arbitrary network: An analytic approach

George W A Constable, Alan J McKane

*Theoretical Physics Division, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK*

---

## Abstract

We analyse a model consisting of a population of individuals which is subdivided into a finite set of demes, each of which has a fixed but differing number of individuals. The individuals can reproduce, die and migrate between the demes according to an arbitrary migration network. They are haploid, with two alleles present in the population; frequency independent selection is also incorporated, where the strength and direction of selection can vary from deme to deme. The system is formulated as an individual-based model, and the diffusion approximation systematically applied to express it as a set of nonlinear coupled stochastic differential equations. These can be made amenable to analysis through the elimination of fast-time variables. The resulting reduced model is analysed in a number of situations, including migration-selection balance leading to a polymorphic equilibrium of the two alleles, and an illustration of how the subdivision of the population can lead to non-trivial behaviour in the case where the network is a simple hub. The method we develop is systematic, may be applied to any network, and agrees well with the results of simulations in all cases studied and across a wide range of parameter values.

**Keywords:** Metapopulation, Migration, Selection, Moran model, Fast-mode reduction

---

## 1. Introduction

The founders of population genetics, who reconciled Mendelian genetics and Darwinian evolution, did so through the use of mathematical models which were frequently deterministic and which involved the key processes of mutation, migration and selection [20]. Crucially, Fisher and Wright added genetic drift to this list by considering simple stochastic processes in systems where the population size,  $N$ , was finite [16, 44]. Subsequent work tended to follow their original approach, assuming discrete generations and discrete state variables (corresponding to the number of individuals in the population carrying one particular type of allele) [12]. However adding more complexity to the models in the form of selection and migration makes this approach, based as it is on Markov chains, very unwieldy, and mathematical progress can be difficult [13].

The solution to this dilemma is to take a mesoscopic perspective. That is, one uses a diffusion approximation in which the (discrete) number of individuals carrying a particular allele,  $n$ , is replaced by the (continuous) fraction  $x = n/N$  carrying that allele [9]. Assuming  $x$  is continuous is usually a good approximation for reasonably large  $N$ . This approximation, although originally suggested by Fisher [15], was popularised by Kimura [8, 23], and proved to be a powerful tool and the starting point for many studies of more complex processes in population genetics [21, 24, 22]. Nevertheless, some quite straightforward models give rise to rather complicated equations even within

this approximation. For instance, a model with migration between  $\mathcal{D}$  subpopulations (demes) which includes selection, leads to a nonlinear partial differential equation in  $\mathcal{D}$  variables for the probability distribution function (pdf) [5] which seems quite intractable.

In this paper we show that equations such as these are in fact not as intractable as they seem, and in many cases can be reduced to a differential equation for a single variable, which can be straightforwardly analysed. The methodology which allows this reduction is based on the elimination of fast variables, and relies on a number of factors. First, rather than formulating the diffusion approximation in terms of a partial differential equation for the pdf, it is more useful to work in terms of an equivalent stochastic differential equation (SDE) [27]. This is a direct generalisation of the equation describing the deterministic dynamics [28]. Second, this formulation of the system dynamics allows us to use much of the same intuition that is used to understand the deterministic process. In particular, we will see that in many of the cases of interest the dynamics can be divided into ‘fast’ variables and ‘slow’ variables [38]. After a short time, the dynamics of the fast variables may be ignored, since they have decayed to their stationary values; all the dynamics is contained in the few (in our case, frequently only one) slow mode. Third, the method is systematic and intuitive, and also applies to more general systems, such as those with many alleles and those involving other processes.

Our intention here is to apply this methodology to the

case of a simple two-allele, haploid model with migration, even though it may be extended to deal with more complicated systems. Population genetics models featuring migration were first considered by Wright [44], who looked at what is now often referred to as the standard island model [36]. Instead of the well-mixed population of size  $N$  which had been previously studied, he considered a set of  $\mathcal{D}$  well-mixed subpopulations. With migratory individuals being chosen from the global population, there was no spatial structure assumed, only interactions between the various subpopulations. These are called demes in the genetic context [20], although within the modern nomenclature of ecology, these are effectively metapopulations [25]. Along with references to islands, we will use the terms demes and metapopulations interchangeably in this paper to refer to areas in which there is no spatial structure, but between which interactions can occur. The case of one deme therefore should reduce to the well mixed case.

Subsequently, the formulation of the stepping stone model [24] introduced what was a very simple topology into the description of migration; the islands were ordered, with migration from island  $i$  only allowed onto islands  $i - 1$  and  $i + 1$ . Maruyama compared selection in the stepping stone model and in the island model [26]. He concluded that if selection was additive (i.e. frequency independent) and local, with the same selection pressure in all demes, then deme population structure played no role. In other words, the population behaved approximately as a well-mixed and spatially homogeneous population, with a size equivalent to the sum of the deme population sizes.

The books by Ewens and Moran [13, 31] describe variants of these models, analyses and conclusions, but for our purposes the next result of note is the work of Nagylaki [32] who studied what would be in modern terminology an arbitrary network of demes. He constructed a migratory model with discrete generations (of the Wright-Fisher type) in the limit of strong migration, that is where the probability of a migration event is of the same order as that of a birth or death event. The effect of this assumption was to create a separation of timescales in the Markov chain. Nagylaki then employed his earlier results on Markov chains with timescale separation [11] to achieve an equation in the diffusion limit. Starting with a neutral model, it was concluded that in the long-time limit the population behaved as if it were well mixed, but with an effective population size less than or equal to the total unstructured population. Equality was shown to be achieved only if the migration matrix was symmetric.

The analysis which is used to reach these conclusions seems, to us at least, difficult to follow, with some parts of the proof relying on results from the theory of Markov chains and others relying in the nature of the diffusion approximation. Nevertheless, the results of the analysis are widely quoted and utilised. The work was extended [32] to the case of different selection strengths on different islands, showing that once again the population was well-mixed with an effective population size, but now also with an

effective selection coefficient. The situation where the selection on different islands operates in different directions was not discussed. In this case within certain parameter ranges a stable fixed point emerges, allowing coexistence of deleterious alleles in some demes, but disadvantageous in others. The deterministic implications of this have been discussed in [31, 14, 33].

In the wake of this work a number of studies were carried out and a plethora of results obtained, all with a variety of different approximations and objectives. Several of these were concerned with an effort to determine the effective population size, which amounts to a rescaling of time for the structured population. Here we will avoid the temptation to describe the results that we obtain in terms of an effective population size, due to its amorphous definition, and at times misleading designation. We refer the reader to [6] for a review of such work. Nagylaki's results on the diffusion limit of Markov chains with a separation of timescale have also been employed in [37], where they were used to good effect to extend the results to systems where fitness is not just additive, but frequency dependent. In this case, however, it becomes important to specify carefully between whom individuals compete (on their own island, on islands connected to their own, or the whole population) as these can sometimes lead to different results. In turn, other work has focused instead on the effect of migration on local deme properties [4].

The approach that we adopt in this paper will be to carefully define the model in terms of individuals (*i.e.* at the microscale). We will work within the context of continuous time Markov chains, that is, in terms of master equations [17]. We will therefore not assume non-overlapping generations as in the Wright-Fisher model, but instead work with the continuous time Moran process [13]. As is well known, these two processes are essentially identical at medium to long times, up to a redefinition of time scales. The master equations for the Moran process involving birth, death, migration, mutation, and so on can be written down in a systematic way [5], although it is too complicated to allow analytic progress to be made. As we have indicated the key to further progress is to write down a mesoscopic description which is achieved through a diffusion approximation which is derived by expanding the master equation in inverse deme size.

A related set of questions to those that we ask here have been studied in a model of language evolution [1, 2, 3], in which each island is mapped on to a speaker having two different linguemes (different ways of saying the same thing) whose concentrations are modified through interaction events (analogous to migration events). While this model has similar features to the one we discuss here, it is distinct, and the methods of analysis and the final results are also different. We have already mentioned the work of Nagylaki [32]. Once again our model, analysis and conclusions differ. Throughout this paper we will stress the systematic way that the underlying individual based model (IBM) can be constructed, and the straightforward and in-

tuitive way that the mesoscopic version of this model can be reduced to an effective theory which can be analysed exactly. All approximations which we make will be checked through the use of numerical simulations in the form of the Gillespie algorithm [19] applied to the underlying IBM.

In Section 2.1 we will introduce the model we use by studying the neutral Moran model on one island. We will illustrate the construction of the master equation and the way in which the mesoscopic description is obtained. We will also give the well-known results for the probability of fixation and mean time to fixation, which will be the quantities of interest in later sections. The generalisation to  $\mathcal{D}$  demes will then be discussed in Section 2.2, and the corresponding mesoscopic description obtained. Similarly in Section 2.3 the one island Moran model with selection will be reviewed, before the general model with migration and selection is described in Section 2.4.

In Section 3, we give a description of the model reduction method in both the neutral case, and the case with selection. The mathematical derivation of the results we use is given in [7]. Here we simply state the one-dimensional reduced equation that approximates the full system. In Section 4 we explore the predictions of the reduced model, calculating the probability of fixation and the mean time to fixation in the neutral case in Section 4.1 and the case with selection in Section 4.2. We find the approximation captures the behaviour of the system remarkably well. We then proceed to apply these results on the probability and time to fixation to two specific systems of interest.

The first, discussed in Section 4.3, is a system in which migration and selection balance to induce a polymorphic equilibrium. To our knowledge such results have only been obtained previously for standard island models [39] and two deme cases with symmetric migration [18], both of which form a subset of the cases we address here. While work by Whitlock [42] allows asymmetric migration and multiple demes, the selection strength may only take on two distinct values in those many demes.

Finally, in Section 4.4, we will give an illustrative example of the predictive power of the reduction by considering the case of a ‘hub’ or ‘spoke’ topology, where a central island is connected to  $(\mathcal{D} - 1)$  other islands, none of which are connected to each other. The subdivision of the population is seen to have a non-trivial effect on the system which the reduced model accurately predicts.

Three appendices are given in which technical details are discussed. The first, Appendix A, covers the Kramers-Moyal expansion of the microscopic model, the second, Appendix B, the general calculation of the probability of fixation and the third, Appendix C, the calculation of the mean time to fixation.

## 2. The migration models

### 2.1. The neutral Moran model and the diffusion limit

In order to make our analysis of the Moran model with coupled migration clear, we first briefly review the well-

mixed Moran model. We begin by formulating this as an IBM with a dynamics given by a master equation, before moving to the continuous limit, in which we arrive at a Fokker-Planck equation (FPE). Readers who wish to see mathematical details of these stages can consult Appendix A. In the main text however we will attempt to restrict attention to the key results and conceptual ideas behind the techniques we will use. Throughout we will make explicit any assumptions made, or relationships inferred in the simplification and analysis of the system.

The population we consider is finite, well-mixed and composed of haploid individuals containing one of two alleles,  $A$  and  $B$ ; the number of each is given by the integers  $n$  and  $m$  respectively. At a specific point in time we pick an individual to reproduce. We assume that the progeny of this reproduction event carries the allele of its parent and that it immediately displaces a pre-existing individual at random. In this way the population size is kept constant so that at any one time  $n + m = N$ , where  $N$  is the total size of the well-mixed population. This is the simplest version of the Moran model [30, 31]. A set of transition rates can then be defined which describe the probability per unit time that allele  $A$  increases or decreases in the population. In this case the transition rates may be obtained from simple combinatorics, and are given by

$$T(n+1|n) = \frac{1}{N(N-1)}n(N-n),$$

$$T(n-1|n) = \frac{1}{N(N-1)}n(N-n).$$

These, together with the master equation [40]

$$\begin{aligned} \frac{dp(n,t)}{dt} &= T(n|n-1)p(n-1,t) \\ &+ T(n|n+1)p(n+1,t) \\ &- [T(n+1|n) + T(n-1|n)]p(n,t), \end{aligned} \quad (1)$$

define the evolution in time of the probability distribution  $p(n,t)$ . The interpretation of the master equation is intuitively clear: the probability that the system is in state  $n$  increases with the probability that the system moves into it from one of the surrounding states,  $n-1$  or  $n+1$ , but decreases with the probability that the system is already in state  $n$  but transitions to another state. Despite this simple description, the master equation is very rarely analytically tractable. We must resort to solving instead an approximation of the equation.

The diffusion approximation leads to a more tractable equation than the full master equation. It involves an expansion in inverse system size, and is valid in the limit of large  $N$ . First we change variables to  $x = n/N$  and Taylor expand the governing master equation in powers of  $N^{-1}$ . Formally this is known as the Kramers-Moyal expansion [17, 35]. Truncating the series at second order

in  $N^{-1}$  leads to the FPE which has the form [28]

$$\begin{aligned} \frac{\partial p(x, t)}{\partial t} = & - \frac{1}{N} \frac{\partial}{\partial x} [A(x)p(x, t)] \\ & + \frac{1}{2N^2} \frac{\partial^2}{\partial x^2} [B(x)p(x, t)], \end{aligned} \quad (2)$$

where the functions  $A(x)$  and  $B(x)$  are given respectively by

$$A(x) = 0, \quad B(x) = 2x(1 - x). \quad (3)$$

It is common practice [31, 13] to rescale time by introducing  $\tau$  such that  $t = N\tau$ , but here we retain the  $N$  dependence for clarity.

A more useful starting point for calculating many quantities of interest is not the FPE itself, but the backward Fokker-Planck equation, which is formally the adjoint of the FPE [17, 35]. In the usual, or forward, FPE we impose initial conditions at time  $t_0$  and find  $p(x, t)$  for  $t > t_0$ . In the backward FPE we impose conditions at time  $t$  and ask what solution  $q(x_0, t_0)$  at an initial time  $t_0 < t$  gave this final condition. For this reason the backward equation is most useful in the investigation of, for instance, the probability of the fixation of allele  $A$  given some initial concentration  $x_0$ , denoted by  $Q(x_0)$ , and the mean time it takes either allele  $A$  or  $B$  to fixate,  $T(x_0)$ . Ordinary differential equations for  $Q(x_0)$  and  $T(x_0)$  can be found from the backward FPE. Details are given in many standard texts: general diffusion problems are described in [17, 35], whilst the methods for the specific case under consideration here are illustrated in [12]. After some calculation [31, 13, 17], one finds the equations

$$\frac{A(x_0)}{N} \frac{dQ(x_0)}{dx_0} + \frac{B(x_0)}{2N^2} \frac{d^2Q(x_0)}{dx_0^2} = 0, \quad (4)$$

$$(5)$$

with boundary conditions  $Q(0) = 0$ ,  $Q(1) = 1$ , and

$$\frac{A(x_0)}{N} \frac{dT(x_0)}{dx_0} + \frac{B(x_0)}{2N^2} \frac{d^2T(x_0)}{dx_0^2} = -1, \quad (6)$$

with boundary conditions  $T(0) = 0$ ,  $T(1) = 0$ . For this neutral case the solutions are [13]

$$Q(x_0) = x_0, \quad (7)$$

$$T(x_0) = -N^2 [(1 - x_0) \ln(1 - x_0) + x_0 \ln(x_0)]. \quad (8)$$

These results are shown graphically in figure 1.

## 2.2. Definition of the neutral metapopulation Moran model

This model consists of a series of  $\mathcal{D}$  demes, on each of which well-mixed populations of fixed and finite size exist. The number of individuals island  $i$  contains is given by  $\beta_i N$ , where  $N$  is some typical island size, and  $\beta_i$  is a scaling factor such that  $\beta_i N$  is an integer. The individuals in the population can carry one of two alleles,  $A$  and  $B$ . An independent deme, unconnected to any others and with a

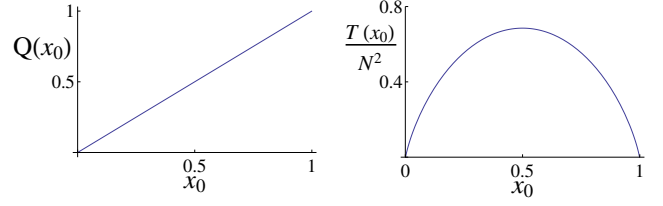


Figure 1: Graphical summary of results from a well-mixed Moran model without selection. Left panel; probability of fixation of allele  $A$ ,  $Q(x_0)$ , in a neutral system as a function of initial  $A$  allele concentration,  $x_0$ . Right panel; time to fixation of either  $A$  or  $B$  allele,  $T(x_0)$ , scaled by the system size squared,  $N^2$ , in a neutral system as a function of the initial concentration of  $A$  allele.

sufficiently large population size, would then be well described by the FPE (2). However we are interested in the form of the FPE for the system which comprises the whole set of  $\mathcal{D}$  demes, with migration between them.

The process is shown diagrammatically in figure 2 for the case  $\mathcal{D} = 2$ . In figure 2(a), a reproduction site is chosen with probability  $f_j$ , which corresponds to a total birth rate for deme  $j$ ; if the demes have an equal birth rate per capita, we simply find  $f_j = \beta_j (\sum_{i=1}^{\mathcal{D}} \beta_i)^{-1}$ . In figure 2(b) either one of the two alleles is chosen to reproduce based on their relative frequencies in that deme. The individual then reproduces and its progeny may either displace an individual in their own deme, or an individual in another deme according to the matrix element  $m_{ij}$  (see figure 2(c)). The matrix  $m_{ij}$  is then the probability that a individual reproducing in  $j$  will have offspring which displaces an individual in  $i$ . Finally, in figure 2(d) the type of individual in  $i$  that is displaced is decided, again based on their relative frequencies in  $i$ . The vector  $f_j$  and the matrix  $m_{ij}$  represent probabilities and so satisfy the conditions  $\sum_j f_j = 1$  and  $\sum_i m_{ij} = 1$  for all  $j$ .

We denote the number of  $A$  alleles in deme  $i$  by  $n_i$ . The number of  $B$  alleles in deme  $i$  is then given by  $\beta_i N - n_i$ , where  $\beta_i N$  is the total population of that deme. The equivalent transition rates for the process depicted in figure 2 can then be calculated using combinatoric arguments. Starting with deme  $i$  and summing over the  $\mathcal{D}$  demes from which an allele can originate, one obtains

$$\begin{aligned} T(n_i + 1 | n_i) &= \sum_{j=1}^{\mathcal{D}} (f_j) \left( \frac{n_j}{\beta_j N} \right) \times \\ &\quad (m_{ij}) \left( \frac{\beta_i N - n_i}{\beta_i N - \delta_{ij}} \right), \\ T(n_i - 1 | n_i) &= \sum_{j=1}^{\mathcal{D}} (f_j) \left( \frac{\beta_j N - n_j}{\beta_j N} \right) \times \\ &\quad (m_{ij}) \left( \frac{n_i}{\beta_i N - \delta_{ij}} \right), \end{aligned}$$

where the dependence of  $T(\mathbf{n} | \mathbf{n}')$  on the elements of  $\mathbf{n}$  that do not change in the transition have been suppressed.

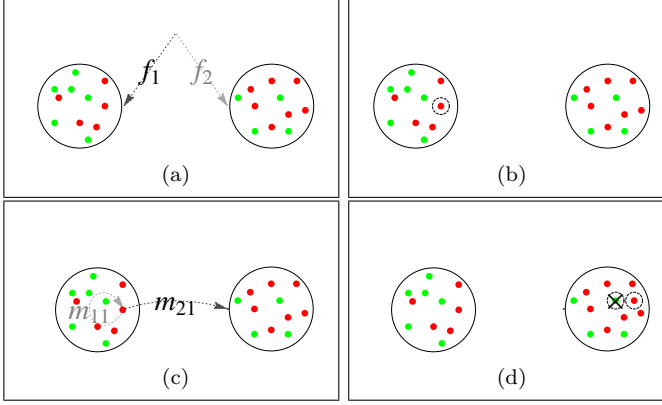


Figure 2: Diagram of the neutral metapopulation Moran model for  $\mathcal{D} = 2$ . Each large black circle is a deme populated by two types of haploid individuals carrying either an allele  $A$ , in red, or allele  $B$  in green. Subfigures (a)-(d) depict the stages in picking an allele to reproduce and an allele to die.

Each of the four factors in these expressions for  $T(n_i \pm 1|n_i)$  corresponds to one of the four processes displayed in figure 2. The diagonal elements of  $m_{ij}$  do not represent a migration process, but instead the probability that an offspring remains in its parent's deme; they are simply equal to one minus the sum of the other elements in the same column to ensure that  $\sum_i m_{ij} = 1$ .

Since  $f_j$  and  $m_{ij}$  always occur together in the combination  $m_{ij}f_j$ , it is convenient to introduce the matrix  $G_{ij} \equiv m_{ij}f_j$ , which we shall call the migration rate matrix, a combination of birth and migratory rates. The migration rate matrix inherits the properties  $\sum_i G_{ij} = f_j$  and  $\sum_{i,j} G_{ij} = 1$ . Again, the diagonal elements of  $G_{ij}$  represent the probability that both deme  $j$  is chosen *and* the progeny of the reproduction remains in the parent deme.

The transition rates in terms of  $G_{ij}$  then become

$$T(n_i + 1|n_i) = G_{ii} \frac{(\beta_i N - n_i)n_i}{(\beta_i N - 1)\beta_i N} + \frac{\beta_i N - n_i}{\beta_i N} \sum_{j \neq i}^{\mathcal{D}} G_{ij} \frac{n_j}{\beta_j N}, \quad (9)$$

$$T(n_i - 1|n_i) = G_{ii} \frac{n_i(\beta_i N - n_i)}{(\beta_i N - 1)\beta_i N} + \frac{n_i}{\beta_i N} \sum_{j \neq i}^{\mathcal{D}} G_{ij} \frac{\beta_j N - n_j}{\beta_j N}, \quad (10)$$

where we have separated out the contribution from the processes involving two islands ( $i$  and  $j$ ) from those which only involve island  $i$ .

The master equation associated with this process (Eq. (A.1) given in Appendix A) is clearly more complicated than that for the well-mixed single-island population, Eq. (1), and no more tractable. Again however, we can make the diffusion approximation. This time we make

the change of variables  $x_i = n_i/(\beta_i N)$  and again expand in powers of  $N^{-1}$  (see Appendix A, with the parameter  $s$  set to zero). Truncating at second order we have the FPE for the metapopulation:

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = - \frac{1}{N} \sum_{i=1}^{\mathcal{D}} \frac{\partial}{\partial x_i} [A_i(\mathbf{x})p(\mathbf{x}, t)] + \frac{1}{2N^2} \sum_{i=1}^{\mathcal{D}} \frac{\partial^2}{\partial x_i^2} [B_{ii}(\mathbf{x})p(\mathbf{x}, t)], \quad (11)$$

with

$$A_i(\mathbf{x}) = \frac{1}{\beta_i} \left( -x_i \sum_{j \neq i}^{\mathcal{D}} G_{ij} + \sum_{j \neq i}^{\mathcal{D}} G_{ij} x_j \right), \quad (12)$$

and

$$B_{ii}(\mathbf{x}) = \frac{1}{\beta_i^2} \left( x_i \sum_{j=1}^{\mathcal{D}} G_{ij} + \sum_{j=1}^{\mathcal{D}} G_{ij} x_j - 2x_i \sum_{j=1}^{\mathcal{D}} G_{ij} x_j \right). \quad (13)$$

It is sometimes assumed that the off-diagonal elements of the matrix  $G_{ij}$  are such that  $G_{ij} = \mathcal{G}_{ij}/N$  for all  $i \neq j$ , where  $\mathcal{G}$  is of order unity [32, 5]. This means that the off-diagonal elements in  $B(\mathbf{x})$  may be neglected, since they give  $\mathcal{O}(N^{-3})$  contributions. Since only the off-diagonal elements of  $G$  appear in the vector  $\mathbf{A}(\mathbf{x})$  and only the diagonal elements of  $G$  appear in  $B(\mathbf{x})$ , both terms on the right-hand side of Eq. (11) are of order  $N^{-2}$ , and they effectively balance each other. Asking that the off-diagonal elements of the matrix  $G$  are small has a clear biological interpretation. The population is strongly subdivided and it is far more likely that an individual's offspring will remain in the deme of its parents than migrate. This is not the generic case however, and the scaling of the off-diagonal terms with the inverse of the population size is in some cases little more than a mathematical convenience.

Here we will make the choice that the elements of the migration matrix are approximately all of the same order. In doing so we are assuming that once a deme is selected, the probability of allele reproduction-migration is not too dissimilar to that of allele reproduction.

### 2.3. The Moran model with selection

To add further complexity to the model prescribed by Section 2.1 one may add the effect of fitness. Specifically we shall incorporate frequency independent fitness, that is, fitness that does not depend on the constitution of the population. The variables  $w_A$  and  $w_B$  are introduced to represent the fitness weightings of alleles  $A$  and  $B$  respectively [34]. The transition rates (see Appendix A) are

then

$$T(n+1|n) = \frac{nw_A}{nw_A + (N-n)w_B} \frac{(N-n)}{N}, \quad (14)$$

$$T(n-1|n) = \frac{(N-n)w_B}{nw_A + (N-n)w_B} \frac{n}{N}. \quad (15)$$

The appearance of  $n$  in the denominator complicates the Kramers-Moyal expansion slightly. This is usually addressed by rewriting the fitness parameters  $w_A = 1 + s$  and  $w_B = 1$ , and expanding in powers of  $s$  under the very reasonable assumption that  $s$  is small. Positive  $s$  indicates a small fitness advantage for individuals with allele  $A$ .

The diffusion approximation leads to the FPE (2), with  $A(x)$  and  $B(x)$  given respectively by

$$A(x) = sx(1-x) \quad \text{and} \quad B(x) = 2x(1-x), \quad (16)$$

for small  $s$ . Here we have included only the lowest order contribution in  $s$  to  $A(x)$ , and omitted the order  $s$  correction in  $B$  altogether, since it will be negligible compared with  $2x(1-x)$  (for further details see Appendix A). In the same manner as in Section 2.1 we find equations Eq. (4) and Eq. (6), with  $A(x)$  and  $B(x)$  given by Eq. (16), for the fixation probability  $Q(x_0)$  and the fixation time  $T(x_0)$  as a function of  $x_0$ , the initial concentration of allele  $A$ .

In this case, solving Eq. (4), the familiar equation for the probability of fixation,  $Q(x_0)$  is found [13]:

$$Q(x_0) = \frac{1 - \exp(-Nsx_0)}{1 - \exp(-Ns)}. \quad (17)$$

While the mean time to fixation can also be obtained analytically (see Eqs. (45) and (46) with  $M$  set to  $N$  and  $\sigma$  set to  $s$ , in Section 4.2), it is sufficient to determine it numerically. Illustrative plots are shown in figure 3.

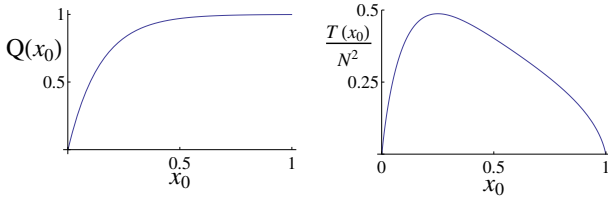


Figure 3: Probability of fixation of allele  $A$ ,  $Q(x_0)$ , and mean time to fixation of either allele,  $T(x_0)$ , in a system where allele  $A$  has a selective advantage  $s$  over allele  $B$ , as a function of initial  $A$  concentration. While  $Q(x_0)$  is obtained from Eq. (17),  $T(x_0)$  has been obtained numerically.

#### 2.4. The Moran model with fitness and migration

We now proceed to incorporate selection into the migration model defined in Section 2.2. We begin as before with a set of transition rates, however now the probability of an individual coming into deme  $i$  from deme  $j$  is a function of the progenitor's fitness in deme  $j$ . In a similar fashion to the case presented in Section 2.3, the fitness of

allele  $A$  on each deme is denoted by the vector  $\mathbf{w}_A$  while the fitness of allele  $B$  is  $\mathbf{w}_B$ . Further details are given in Appendix A, though the progression from the one deme model to the  $\mathcal{D}$  deme model is straightforward. The transition rates are

$$\begin{aligned} T(n_i+1|n_i) &= \sum_{j=1}^{\mathcal{D}} \frac{(\beta_i N - n_i)}{\beta_i N - \delta_{ij}} G_{ij} \frac{[\mathbf{w}_A]_j n_j}{[\mathbf{w}_A]_j n_j + [\mathbf{w}_B]_j (\beta_j N - n_j)}, \\ T(n_i-1|n_i) &= \sum_{j=1}^{\mathcal{D}} \frac{n_i}{\beta_i N - \delta_{ij}} G_{ij} \frac{[\mathbf{w}_B]_j (\beta_j N - n_j)}{[\mathbf{w}_A]_j n_j + [\mathbf{w}_B]_j (\beta_j N - n_j)}. \end{aligned} \quad (18)$$

Letting  $[\mathbf{w}_B]_i = 1$  for every island  $i$ , the elements of the fitness term  $[\mathbf{w}_A]$  are now dependent on both the typical selection strength,  $s$ , and the vector  $\boldsymbol{\alpha}$ , which moderates the typical selection strength in magnitude and direction such that

$$[\mathbf{w}_A]_i = 1 + s\alpha_i. \quad (19)$$

Positive  $\alpha_i$  therefore corresponds to allele  $A$  being advantageous relative to  $B$  on island  $i$ , while negative  $\alpha_i$  means  $A$  is deleterious. We assume that the elements of  $\boldsymbol{\alpha}$  are of order unity.

Proceeding in the in the same spirit as Section 2.2, we conduct a Kramers-Moyal expansion to arrive at Fokker-Planck equation Eq. (11) valid in the limit of large  $N$  and small  $s$ . The calculation is carried out in full in Appendix A. The FPE is defined through an  $\mathbf{A}(\mathbf{x})$  vector and a diagonal  $\mathbf{B}(\mathbf{x})$  matrix which, when expressed as a series in  $s$ , have elements

$$\begin{aligned} A_i(\mathbf{x}) &= \frac{1}{\beta_i} \left\{ \sum_{j \neq i}^{\mathcal{D}} G_{ij} (x_j - x_i) + s \sum_{j=1}^{\mathcal{D}} G_{ij} \alpha_j x_j (1 - x_j) \right. \\ &\quad \left. - s^2 \sum_{j=1}^{\mathcal{D}} G_{ij} \alpha_j^2 x_j^2 (1 - x_j) \right\} + \mathcal{O}(s^3). \end{aligned} \quad (20)$$

and

$$\begin{aligned} B_{ii}(\mathbf{x}) &= \frac{1}{\beta_i^2} \left\{ x_i \sum_{j=1}^{\mathcal{D}} G_{ij} + \sum_{j=1}^{\mathcal{D}} G_{ij} x_j \right. \\ &\quad \left. - 2x_i \sum_{j=1}^{\mathcal{D}} G_{ij} x_j \right\} + \mathcal{O}(s). \end{aligned} \quad (21)$$

### 3. The approximation procedure

Having derived the FPEs for a sequence of progressively more complex situations, we can now describe the approximation which is the main subject of this paper — the reduction of the full model to an effective one-dimensional system. The full details of the approximation method and calculation are given in [7]. Here we will give

a description of the technique before stating the equation of the reduced system, and then applying it.

The reduction technique relies on utilising concepts from deterministic dynamical systems theory in order to understand and simplify the stochastic equations. For this reason it is more convenient to work in the context of SDEs, rather than the FPE. In general, the system described by the FPE (11), with  $\mathbf{A}(\mathbf{x})$  given by Eq. (20) and  $B(\mathbf{x})$  by Eq. (21), is entirely equivalent to the Itô SDE [17, 35]

$$\dot{x}_i = A_i(\mathbf{x}) + \frac{1}{\sqrt{N}}\eta_i(\tau), \quad (22)$$

where the dot indicates differentiation with respect to  $\tau = t/N$  and  $\boldsymbol{\eta}(\tau)$  is a Gaussian white noise with zero mean and correlation functions

$$\langle \eta_i(\tau)\eta_j(\tau') \rangle = B_{ij}(\mathbf{x})\delta(\tau - \tau'). \quad (23)$$

We may intuitively think of this as the deterministic system,  $\dot{x}_i = A_i(\mathbf{x})$ , with a small amount of added noise. Before proceeding it is useful to re-write  $A_i(\mathbf{x})$ , highlighting its linearity at zeroth order in  $s$ :

$$A_i(\mathbf{x}) = \sum_{j=1}^{\mathcal{D}} H_{ij}x_j + \beta_i^{-1} \left\{ s \sum_{j=1}^{\mathcal{D}} G_{ij}\alpha_j x_j(1-x_j) - s^2 \sum_{j=1}^{\mathcal{D}} G_{ij}\alpha_j^2 x_j^2(1-x_j) \right\}, \quad (24)$$

where the matrix  $H$  has elements

$$H_{ij} = \frac{G_{ij}}{\beta_i} \text{ if } i \neq j, \quad H_{ii} = -\sum_{j \neq i}^{\mathcal{D}} \frac{G_{ij}}{\beta_i}. \quad (25)$$

The neutral system is then clearly obtained by setting  $s = 0$ .

The properties of the matrix  $H$  are central to the application of our method, and so we briefly summarise them. In general  $H$  will not be symmetric, and so its eigenvalues may be complex, and it will have distinct right- and left-eigenvectors. We will denote the right- and left-eigenvectors corresponding to the eigenvalue  $\lambda^{(i)}$  as  $\mathbf{v}^{(i)}$  and  $\mathbf{u}^{(i)}$  respectively. They are orthogonal if they correspond to different eigenvalues and can be defined so that they are orthonormal:

$$\sum_{k=1}^{\mathcal{D}} v_k^{(i)} u_k^{(j)} = \delta_{ij}. \quad (26)$$

Furthermore, the first eigenvalue of  $H$ ,  $\lambda^{(1)}$ , is zero, and the first right eigenvector has components  $v_i^{(1)} = 1$  for all  $i$  and for any choice of parameters. All other eigenvalues of  $H$  can be shown to have a negative real part [7].

Firstly we consider the neutral deterministic system, that is, we set  $s = 0$  and  $N \rightarrow \infty$ . The system is then linear, entirely governed by the matrix  $H$  and hence exactly

solvable. The deterministic solution displays a distinct separation of timescales; the system quickly collapses onto a linear subspace, the centre manifold [43], upon which it remains indefinitely. It is this separation of timescales which we exploit in order to make analytic progress. If one is only interested in the long-term behaviour of the system, one may effectively ignore this rapid transient and assume that the system reaches the centre manifold very quickly. One can show that this centre manifold is given by  $x_i = x_j$ , for all  $i, j$  (parallel to the vector  $\mathbf{v}^{(1)}$ ) independent of the choice of parameters (see figure 4, upper panel, blue-dashed line).

Let us continue to consider the neutral system, but now allow the population to be finite so that noise is non-zero. Away from the centre manifold we find that the deterministic dynamics dominate the system's trajectory, dragging it to the vicinity of the centre manifold. Once in this region, the deterministic dynamics cease to dominate and the effect of demographic noise comes into play. This situation is depicted in the upper panel of figure 4. The deterministic dynamics keep the trajectory of the system confined to a region around the centre manifold, along which it moves stochastically.

We introduce a variable  $z$  which represents the component of the state vector along the centre manifold. The variable  $z$  lies on the interval  $[0, 1]$ , so that if  $z = 0$  then  $x_i = 0$  for all  $i$ , and likewise if  $z = 1$ ,  $x_i = 1$  for all  $i$ . In order to exploit the separation of timescales in this stochastic setting, we make the following assumptions:

- (i) The system lies on the centre manifold.
- (ii) The only component of the deterministic dynamics,  $\mathbf{A}(\mathbf{x})$ , is along the centre manifold.
- (iii) The only component of the noise is in the direction of the centre manifold.
- (iv) Given a set of initial conditions  $\mathbf{x}_0$ , the system quickly collapses along a deterministic trajectory to some point,  $z_0$ , on the centre manifold.

This allows us to construct a one-dimensional SDE that describes the dynamics of the system in terms of  $z$ .

Now what if the system features selection? Firstly the deterministic analogue of the SDE (22) is now non-linear. However, if  $s$  is small (which has already been assumed in the Taylor expansion of Eq. (18)) one still observes a separation of timescales. The system now quickly collapses to a curved slow subspace rather than a centre manifold. An analytic approximation to the slow subspace is still obtainable, by using a linear approximation for the directions of the transient behaviour. This approximation provides excellent agreement with the slow-subspace observed by numerically simulating the deterministic ODE's, as demonstrated in the bottom panel of figure 4. As  $s \rightarrow 0$  the centre manifold and slow subspace coincide.

In order to extend the stochastic reduction method to the case with selection, we alter our assumptions slightly, relying in part on the small size of  $s$  relative to  $H$ . Maintaining our definition of  $z$  as measuring the distance along



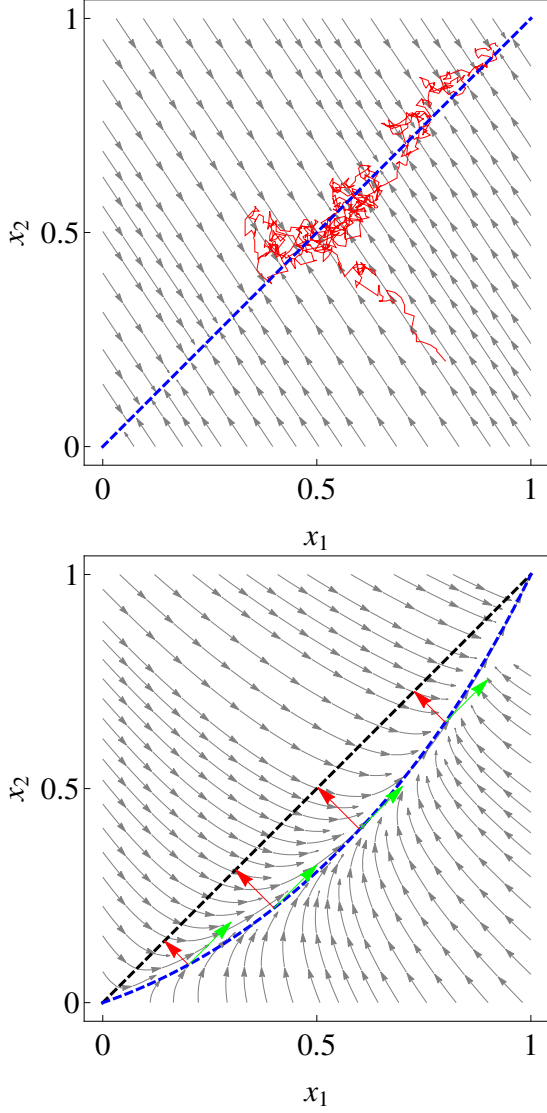


Figure 4: Top panel: Time series of an individual stochastic trajectory (red) and deterministic trajectories (grey) for the neutral system,  $s = 0$ . The stochastic trajectory can be seen to quickly collapse onto the deterministic centre manifold  $x_1 = x_2$ , highlighted in blue, along which stochastic dynamics are observed. Bottom panel: Deterministic trajectories for  $s \neq 0$  plotted in grey. Here a large value of  $s$  has been used ( $s = 0.1$ ), in order to emphasise the nature of the dynamics. The order  $s$  and  $s^2$  non-linear terms in Eq. (20) result in trajectories that are curved relative to the neutral case depicted in the top panel. The variable  $z$  is measured along the straight black dashed line, while the approximation of the slow subspace is plotted as a curved blue line. The distance between the slow subspace and the line  $\mathbf{x} = z\mathbf{v}^{(1)}$  is a function of  $s$ . The key elements of the approximation are that the system lies on this slow subspace, can move in the direction  $\mathbf{v}^{(1)}$  (green arrows), but cannot move along the directions  $\mathbf{v}^{(2)}$  (red arrows).

$\mathbf{v}^{(1)}$  (the centre manifold in the neutral case), we make the following approximations:

- (i) The system lies on the slow subspace.
- (ii) The only component of the deterministic dynamics is in the direction  $\mathbf{v}^{(1)}$ .
- (iii) The only component of the noise is in the direction  $\mathbf{v}^{(1)}$ .
- (iv) Given a set of initial conditions  $\mathbf{x}_0$ , the system quickly collapses to a point on the slow subspace, whose component in the direction  $\mathbf{v}^{(1)}$  is approximately that of the neutral case,  $z_0$ .

This situation is depicted in the bottom panel of figure 4; the system moves freely in the direction  $\mathbf{v}^{(1)}$ , indicated by green arrows, but does not move in the direction  $\mathbf{v}^{(2)}$ , parallel to the red arrows. Again we arrive at a one-dimensional SDE approximating the dynamics along the slow subspace.

The general reduced SDE may be expressed in terms of the variable  $z$  as

$$\dot{z} = \bar{A}(z) + \frac{1}{\sqrt{N}}\zeta(\tau), \quad (27)$$

with

$$\langle \zeta(\tau)\zeta(\tau') \rangle = \bar{B}(z)\delta(\tau - \tau'). \quad (28)$$

The reduced deterministic term is given by [7]

$$\begin{aligned} \bar{A}(z) = & sa_1z(1-z) + s^2a_2z^2(1-z) \\ & + s^2a_3z(1-z)(1-2z) + \mathcal{O}(s^3), \end{aligned} \quad (29)$$

where parameters  $a_1$  and  $a_2$  are defined by

$$a_1 = \sum_{i,j=1}^{\mathcal{D}} u_i^{(1)} \frac{G_{ij}}{\beta_i} \alpha_j \quad (30)$$

and

$$a_2 = - \sum_{i,j=1}^{\mathcal{D}} u_i^{(1)} \frac{G_{ij}}{\beta_i} \alpha_j^2, \quad (31)$$

and the parameter  $a_3$ , whose more complicated form arises from contributions on the slow subspace, is

$$a_3 = - \sum_{i,j,k,l=1}^{\mathcal{D}} u_i^{(1)} \frac{G_{ij}}{\beta_i} \alpha_j \left( \sum_{a=1}^{\mathcal{D}-1} \frac{v_j^{a+1} u_k^{(a+1)}}{\lambda^{(a+1)}} \right) \frac{G_{kl}}{\beta_k} \alpha_l. \quad (32)$$

The noise covariance matrix  $\bar{B}(z)$ , takes the relatively simple form

$$\bar{B}(z) = 2b_1z(1-z) + \mathcal{O}(s), \quad (33)$$

where we have introduced the parameter

$$b_1 = \sum_{i,j=1}^{\mathcal{D}} [u_i^{(1)}]^2 G_{ij} \beta_i^{-2}. \quad (34)$$



Finally, the initial condition for the one-dimensional system must be defined mathematically. Isolating the component of the initial state vector,  $\mathbf{x}_0$ , along the centre manifold [7], one finds

$$z_0 = \sum_{i=1}^{\mathcal{D}} u_i^{(1)} x_{0i}. \quad (35)$$

These expressions completely define the reduced system.

Our approximate system effectively ignores initial transient dynamics under the assumption that these are negligible when asking questions about the long-time properties of the system. The reduced system is therefore ideally suited to answering questions about fixation, which will occur on a much slower timescale than it takes for the system to reach the centre manifold or slow subspace. We expect that the reduced system will provide a good approximation to the full system so long as the linear deterministic dynamics are dominant. This has been found to be the case so long as the magnitudes of the real parts of the non-zero eigenvalues of  $H$  are greater than  $N^{-1/2}$  and  $s$ , as discussed in [7].

#### 4. Analysing the reduced model

It has been stated in the previous section that, given the nature of the fast-timescale approximation, the reduced system is suited to answering questions about fixation. In order to test the predictions of the reduced model against stochastic simulations of the full model defined by Eq. (18), we choose to use the probability of fixation and time to fixation as metrics.

We begin by recalling that the Itô SDE Eq. (27) is entirely equivalent to the FPE

$$\begin{aligned} \frac{\partial p(z, t)}{\partial t} = & - \frac{1}{N} \frac{\partial}{\partial z} [\bar{A}(z)p(z, t)] \\ & + \frac{1}{2N^2} \frac{\partial^2}{\partial z^2} [\bar{B}(z)p(z, t)]. \end{aligned} \quad (36)$$

We can therefore calculate, just as in the well-mixed case discussed in Section 2.1, the probability of fixation,  $Q(z_0)$  and the time to fixation  $T(z_0)$  by solving Eq. (4) and Eq. (6) with  $x$  replaced by  $z$  and using the reduced terms  $\bar{A}(z)$  and  $\bar{B}(z)$  given by Eq. (29) and Eq. (33).

##### 4.1. Fixation probability and mean fixation time in the neutral case

To obtain the neutral reduced model we simply set  $s = 0$  in Eq. (27), which yields  $\bar{A}(z) = 0$  by Eq. (29), with  $\bar{B}(z)$  given by Eq. (33). The reduced equation has the same functional form as that of the well-mixed case, Eq. (3), but with  $N^2$  scaled by  $b_1$ . Thus the calculations of the fixation probability,  $Q(z_0)$ , and fixation time,  $T(z_0)$ , follow in a straightforward manner:

$$Q(z_0) = z_0, \quad (37)$$

$$T(z_0) = -(N_{\text{Tot}} r_N)^2 [(1 - z_0) \ln(1 - z_0) + z_0 \ln(z_0)], \quad (38)$$

where

$$N_{\text{Tot}} = N \sum_{k=1}^{\mathcal{D}} \beta_k \quad \text{and} \quad r_N = \left( \sqrt{b_1} \sum_{k=1}^{\mathcal{D}} \beta_k \right)^{-1}. \quad (39)$$

These are identical to the results in the well-mixed case if in Eqs. (7) and (8) we replace  $x_0$  by  $z_0$  and replace  $N$  by  $N_{\text{Tot}} r_N$ .

The parameter  $N_{\text{Tot}}$  is the total size of the unstructured population. A natural interpretation is that the population behaves as a well-mixed population with a new ‘effective population size’  $r_N N_{\text{Tot}}$ . We will however avoid this terminology, since the variable  $z$  is not directly equivalent to the allele frequency within the global population and, in addition, since it is frequently used in other situations in which its meaning differs from that which we would ascribe to it [41, 13].

In agreement with the results of Nagylaki [32] (whom we recall considered a Wright-Fisher migration model with non-overlapping generations), the value of  $r_N$  takes rather a simple form in the case that migration rate matrix,  $G$ , is symmetric [7]. Direct substitution using Eq. (25) and the symmetry of  $G$  shows that  $\sum_{i=1}^{\mathcal{D}} \beta_i H_{ij} = 0$  for all  $j$ , and so  $\beta_i$  is the left-eigenvector of  $H$  with eigenvalue  $\lambda_1 = 0$ . So  $\beta_i$  must be proportional to  $u_i$ , and since the normalisation of  $u_i$  has already been fixed through Eq. (26) and the choice of  $\mathbf{v}^{(1)}$ , we have that  $u_i^{(1)} = \beta_i (\sum_{k=1}^{\mathcal{D}} \beta_k)^{-1}$ . Substituting this expression into Eq. (34) and using  $\sum_{i,j} G_{ij} = 1$ , gives  $r_N = 1$ . When the migration is symmetric therefore, the population behaves on the same timescale as a well-mixed population of equal size, albeit with a weighted initial allele frequency,  $z_0$ .

Our results diverge from those of Nagylaki outside this limit however. In his model and analysis, it was found that  $r_N \leq 1$ , whereas we find no strict upper bound on the value of  $r_N$ . Indeed, in Section 4.4 we will find that  $r_N$  may be significantly higher than one in some particular situations. This serves to emphasise that  $r_N N_{\text{Tot}}$  does not provide an effective population size, but rather describes a typical timescale for fixation.

To demonstrate the range of values  $r_N$  can take, a numerical study can be conducted. An ensemble of random migration matrices,  $m$ , are first generated. We have to be careful to pick the elements of  $m$  such that the normalisation condition  $\sum_{i=1}^{\mathcal{D}} m_{ij} = 1$  holds. Additionally, from a modelling perspective, we would like to see the diagonal elements of  $m$  larger than  $1/2$  at least,  $m_{ii} > 1/2$ , so that the probability of an offspring not migrating is greater than the probability it migrates. For each random migration matrix generated, an  $r_N$  may then be calculated to give an indication of a potential distribution of  $r_N$  values. Since we expect the reduction technique to become unreliable if any of the real parts of the non-zero eigenvalues of  $H$  are smaller in magnitude than  $N^{-1/2}$  (see Section 3 and [7]), we discard any  $m$  matrices that yield such values.

Initially we consider systems with  $\mathcal{D} = 4$  and  $\beta_i = 1$  for all  $i$ . The values for  $r_N$  are plotted in a histogram

in figure 5; while no strict upper value for  $r_N$  exists, the distribution in this parameter regime does not allow for  $r_N > 1$ . We note however, that this is a feature of our modelling choice; if we remove the restriction  $m_{ii} > 1/2$ ,  $r_N$  can take a range of values around one. Further, if we allow the island sizes to vary (as in figure 5, inset) the distribution of  $r_N$  values is altered to allow  $r_N > 1$ .

We can test these predictions against a Gillespie simulation of the full system, defined by the transition rates Eq. (9) and Eq. (10). The time to extinction of one or other of the alleles,  $T(z_0)$ , calculated from the effective theory, Eq. (38), is used as a measure of the timescale of the effective system. We find excellent agreement between simulation and theory across a range of parameters, as shown in the bottom panel of figure 5.

#### 4.2. Fixation probability and mean fixation time in the case with selection

Once again we seek to solve Eq. (4) and Eq. (6) with  $x$  replaced by  $z$  and  $\bar{A}(z)$  and  $\bar{B}(z)$  given by Eq. (29) and Eq. (33), but now with some selective bias for one or other of the alleles, such that  $s \neq 0$ . We begin by noting that Eq. (29) can be written more compactly as

$$\bar{A}(z_0) = sz_0(1 - z_0)(k_1 - sk_2z_0), \quad (40)$$

with

$$k_1 = a_1 + sa_3 \quad \text{and} \quad k_2 = 2a_3 - a_2. \quad (41)$$

We can now solve Eq. (4) to obtain an expression for the probability of fixation. We shall merely state the result here; full details of the calculation are given in Appendix B. Defining the function

$$l(z_0) = \sqrt{\frac{N}{(2b_1|k_2|)}}(sk_2z_0 - k_1), \quad (42)$$

the probability of fixation, given initial weighted frequency of  $A$  allele  $z_0$ , is given by

$$Q(z_0) = \frac{1 - \chi(z_0)}{1 - \chi(1)}; \quad \chi(z_0) = \frac{f(l(z_0))}{f(l(0))}, \quad (43)$$

where the form of the function  $f$  depends on the sign of  $k_2$ . If  $k_2 < 0$  the function  $f$  is the complementary error function [29], if  $k_2 > 0$ , it is the imaginary error function [10].

The form of  $Q(z_0)$  is more complex as compared to the neutral case, for which we found that the metapopulation model behaved analogously to the well-mixed model (see Eq. (37) and Eq. (7)). However, we can gain further insight into the model by considering broadly two different parameter regimes. First, we examine the situation in which the advantageous allele is the same on each of the demes. In this case we find that the parameters  $a_1$ ,  $a_2$  and  $b_1$  are all of order 1, while we have taken  $N$  large. Given that the function  $l(z)$  is then relatively large, we can perform an asymptotic expansion of the error function (see

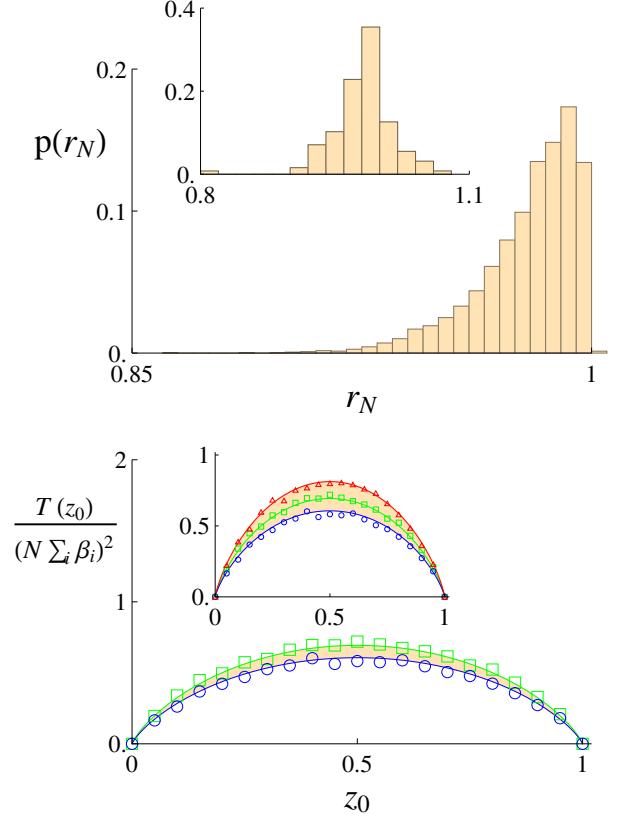


Figure 5: Top panel: Histograms of values of  $r_N$  obtained for a  $\mathcal{D} = 4$  system with randomly generated and appropriately normalised migration matrices,  $m$ , and  $s = 0$ . Main histogram shows results obtained when all island sizes are the same,  $\beta_i = 1$  for all  $i$ . The inset histogram is obtained when the elements  $\beta_i$  are themselves random integers. We have taken  $N = 250$  in both cases, and hence have discounted any of the random systems that yield a non-zero eigenvalues with a real part greater than  $-N^{-1/2}$ . Bottom panel: Plots of the mean time to fixation as a function of the initial condition  $z_0$ . Analytic predictions from the reduced model, Eq. (38), are plotted as continuous lines, while the results from simulation are plotted as symbols. Plots in blue/circles correspond to the smallest  $r_N$  values obtained in the histograms (top panel), those in green/squares are obtained from systems with symmetric  $G$  matrices ( $r_N = 1$ ), and those in red/triangles correspond to the largest  $r_N$  values obtained in the histograms. Once again in the main graph all islands are of the same size, while in the inset plot  $\beta_i$  is allowed to vary.

Appendix B, Eq. (B.10) and Eq. (B.11)) to find that for both  $k_2 < 0$  and  $k_2 > 0$ ,

$$\chi(z_0) \approx \left(1 - \frac{k_2}{k_1}sz_0\right)^{-1} \exp\left(-\frac{k_1}{b_1}sNz_0 + \frac{k_2}{2b_1}s^2Nz_0^2\right).$$

Having obtained this expression, valid for large  $l(z_0)$ , we can make a further approximation for small  $s$ . Taking only

linear  $s$  terms from the above equation, we obtain

$$\chi(z_0) \approx \exp\left(-\frac{a_1}{b_1}sNz_0\right), \quad (44)$$

which is the form given in Eq. (17), but with a selection strength (or system size) weighted by the ratio  $a_1/b_1$ . One may also obtain this solution by simply truncating the reduced term Eq. (29) at order  $s$  rather than  $s^2$ , and solving the relevant ODE, Eq. (4) in Section 2.1. We find that this provides a very good approximation in this regime, as demonstrated in figure 6.

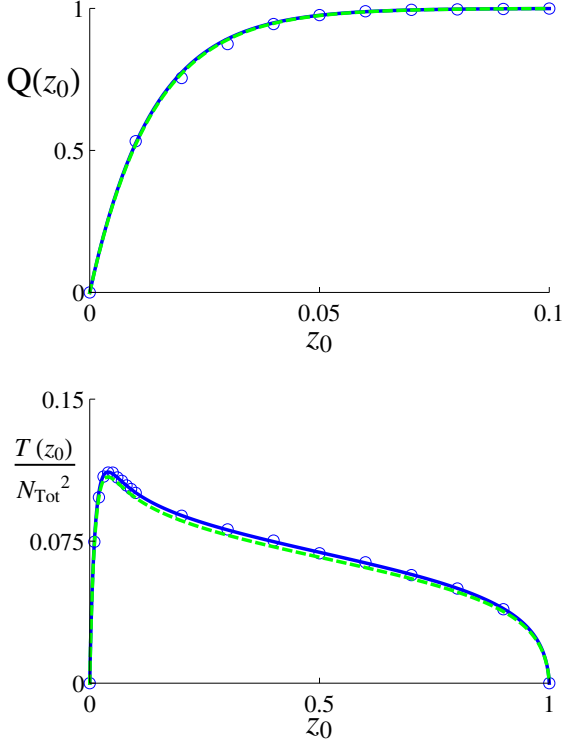


Figure 6: Plots for the probability of fixation,  $Q(z_0)$ , at low values of  $z_0$ , and the mean time to fixation,  $T(z_0)$ , in a system where  $s$  is of order  $N^{-1/2}$ . Continuous blue lines are obtained from the reduced model (using Eq. (43) with  $k_2 > 0$  for  $Q(z_0)$  and solving Eq. (6) with  $\bar{A}(z)$  given by Eq. (29) and  $\bar{B}(z)$  by Eq. (33), numerically for  $T(z_0)$ ). Parameters used here are  $\mathcal{D} = 4$ ,  $s = 0.05$ ,  $N = 400$ ,  $\beta = (1, 1, 1, 2)$ ,  $\alpha = (1, 0.1, 0.5, 1)$ . We omit the explicit form of the migration matrix here for brevity. Since there are no demes in which selection acts in a contrary direction to any of the others, we can use the asymptotic expansion for  $Q(z)$ , Eq. (43) with Eq. (44). The asymptotic expression is plotted by a green dashed line; it is indistinguishable from the full order  $s^2$  solution in this regime. For  $T(z_0)$  we also plot the first order in  $s$  solution as a green dashed line; while qualitatively similar to the full solution, there is some numerical discrepancy.

If the direction of selection varies from deme to deme however, then from a consideration of the forms of  $a_1$  and

$a_2$  one can see that there may be some cancellations. This reduces the size of these parameters and invalidates the use of the asymptotic expansion; one must therefore resort to evaluating the expressions given in Eq. (43) numerically. We find excellent agreement across a wide range of parameters, as shown in [7].

One may also calculate the mean time to fixation,  $T(z_0)$  from Eq. (6) with  $x_0$  replaced by  $z_0$ . There are singular points of the differential equation at the boundaries, and care is required when imposing boundary conditions. These aspects are discussed in Appendix C, where we find expressions for  $T(z_0)$  in terms of well-defined integrals at various order of  $s$ . For instance, to first order in  $s$  we find that

$$T(z_0) = c_2 [1 - e^{-M\sigma z_0}] - M^2 e^{-M\sigma z_0} \int_0^{z_0} dx e^{M\sigma x} [\ln x - \ln(1-x)], \quad (45)$$

where

$$c_2 = \frac{M^2 e^{-M\sigma}}{1 - e^{-M\sigma}} \int_0^1 dx e^{M\sigma x} [\ln x - \ln(1-x)]. \quad (46)$$

Here  $M = N/\sqrt{b_1}$  and  $\sigma = a_1 s/\sqrt{b_1}$ . The integrals in Eqs. (45) and (46) may be expressed as combinations of the exponential integral function [29] and logarithms, but they may also be easily evaluated numerically. At second order in  $s$  the results are more complex, but can again be straightforwardly evaluated. Once again we find very good agreement between the reduced model and simulation (see figure 6 and [7]).

We now proceed to discuss how some of the predictions of the reduced system relate to the behaviour of the full system. In Section 4.3 we will begin by discussing a case in which the system admits a polymorphic equilibrium in the deterministic limit. We will then consider the case of a stylised network topology, the hub, in order to both illustrate the details of the method and to reveal how the partitioning of the population into demes can significantly alter the behaviour of the system.

#### 4.3. Migration-selection balance

So far we have introduced our migration model and applied the reduction technique to arrive at a one-dimensional SDE or FPE which captures the dynamics of the metapopulation. Having obtained these general results, it is now both interesting and instructive to consider a specific system. Of particular note, is the case in which the deterministic system,  $N \rightarrow \infty$ , predicts a migration-selection balance of the two alleles; both alleles  $A$  and  $B$  can coexist in a stable polymorphic equilibrium.

Let us begin by considering the deterministic equations, Eq. (22) with  $\mathbf{A}(x)$  given by Eq. (24) and  $N \rightarrow \infty$ . For clarity we restrict our attention to a two island system

with equal island sizes,  $\mathcal{D} = 2$ ,  $\beta = (1, 1)$ , and a symmetric migration matrix

$$m = \begin{pmatrix} \theta & (1-\theta) \\ (1-\theta) & \theta \end{pmatrix}, \quad (47)$$

parametrised by the appropriately normalised probability of offspring *not* migrating,  $0 < \theta < 1$ . While the behaviour of the linear neutral system was straightforward, the introduction of the non-linear  $s$  terms in Eq. (29) allows for more complicated behaviour. One finds that for  $s \neq 0$ , nine fixed points emerge. Two of these are at the points of fixation  $\mathbf{x}_A^* = (1, 1)$  and  $\mathbf{x}_B^* = (0, 0)$ . While a numerical analysis finds that six of the remaining seven fixed points have values outside the physical range, a final fixed point,  $\mathbf{x}_{PE}^*$ , may arise in between  $\mathbf{x} = (0, 0)$  and  $\mathbf{x} = (1, 1)$ , under the condition that the selective pressure works in opposite directions on each of the demes. Further, one can observe that only one of the fixed points  $\mathbf{x}_A^*$ ,  $\mathbf{x}_B^*$ , or  $\mathbf{x}_{PE}^*$  is stable for a given set of parameters. An overview of the situation is given in figure 7 (top panel); while the region of stable polymorphic equilibrium may appear large in this highly symmetric parameter regime, we note that in general it only occurs for a very restricted range of parameters.

Let us now restrict our attention to a perfectly symmetric set of parameters by setting  $\alpha = (1, -1)$ . A phase diagram for this system is shown in figure 7 (bottom panel). It is interesting to note the position at which the fixed point  $\mathbf{x}_{PE}^*$  is found. One might expect, given the highly symmetric nature of the system, that it would be found equidistant between the points of fixation of allele  $A$  and allele  $B$ ,  $\mathbf{x}_A^*$  and  $\mathbf{x}_B^*$ . While this is true at first order in  $s$ , at second order  $\mathbf{x}_{PE}^*$  is shifted closer to  $\mathbf{x}_B^*$ . Further, the stability of this fixed point increases with increasing  $s$ . We may now ask, how does this deterministic behaviour in such a regime impact the predictions of the reduced stochastic system, Eq. (27)?

Firstly we note that to first order in  $s$ , our deterministic term  $\bar{A}(z)$  in Eq. (29), admits no fixed point other than  $z = 0$  and  $z = 1$ . We would expect, however, that the first order in  $s$  description would work well for particularly small values of  $s$ , say  $s \approx 1/N$ . Indeed, this is what we find for small  $s$ ; the deterministic drive towards the polymorphic fixed point is sufficiently weak that its existence has little effect on the probability of fixation or mean time to fixation. The probability of fixation is then well approximated by Eq. (44) and the time to fixation by Eq. (6) with Eq. (29) to first order in  $s$ , as seen in figure 8 (green/square plot).

For larger values of  $s$ , the stability of the polymorphic fixed point increases in the deterministic limit. To capture the effect on  $Q(z_0)$  and  $T(z_0)$  one must solve Eq. (4) and Eq. (6) to second order in  $s$  (using Eq. (29) in full). One finds the probability of fixation begins to ‘plateau’ across a range of initial conditions as  $s$  increases, with the fixation of allele  $B$  becoming increasingly likely. This counter-intuitive break in symmetry can be viewed as a consequence of the skewed fixed point, which biases the

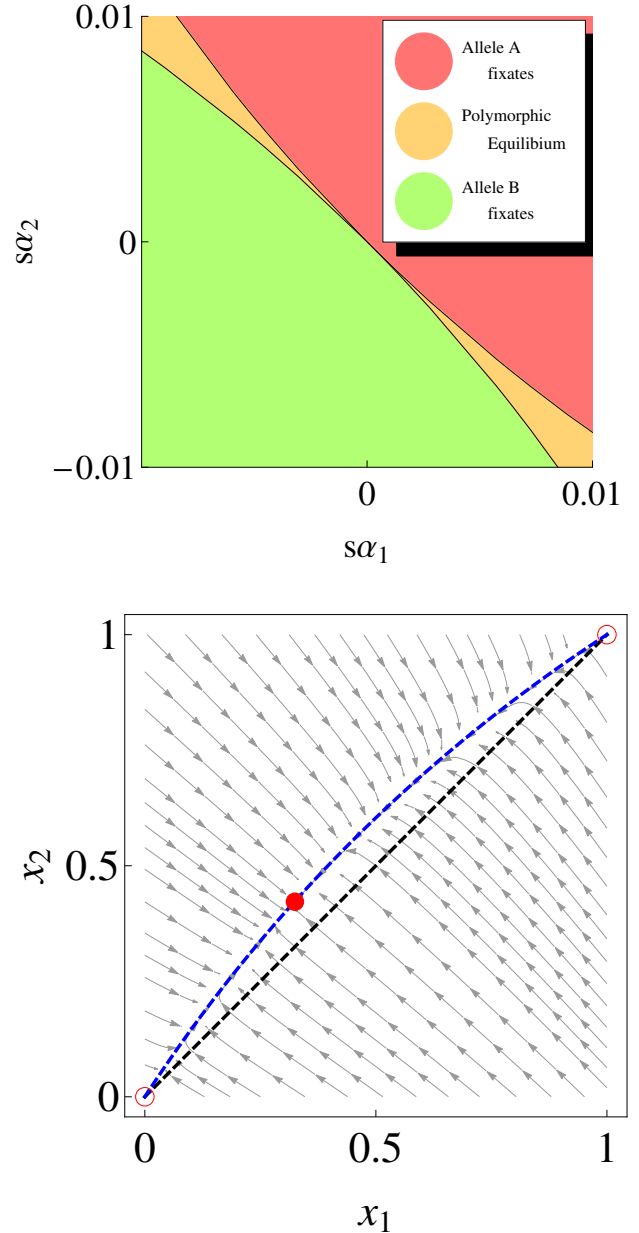


Figure 7: Top panel: Plot of regions of stability for the fixed points  $\mathbf{x}_A^*$  (red, upper-right region),  $\mathbf{x}_B^*$  (green, lower-left region) and  $\mathbf{x}_{PE}^*$  (orange, central region), for a system  $\mathcal{D} = 2$ ,  $\beta = (1, 1)$  and  $m$  given by Eq. (47) with  $\theta = 0.95$ . Bottom panel: Deterministic trajectories (grey) for the same system with  $\theta = 0.8$ ,  $\alpha = (1, -1)$  and  $s = 0.14$ . The stable fixed point  $\mathbf{x}_{PE}^*$  is indicated by a red disc, while unstable fixed points,  $\mathbf{x}_A^*$  and  $\mathbf{x}_B^*$ , are red circles. The straight line  $\mathbf{x} = z\mathbf{v}^{(1)}$  is plotted as a black dashed line, while the analytic approximation of the curved slow subspace is plotted as a blue dashed line. The location of  $\mathbf{x}_{PE}^*$  directly on the approximate slow subspace serves to further emphasise the quality of the approximation.

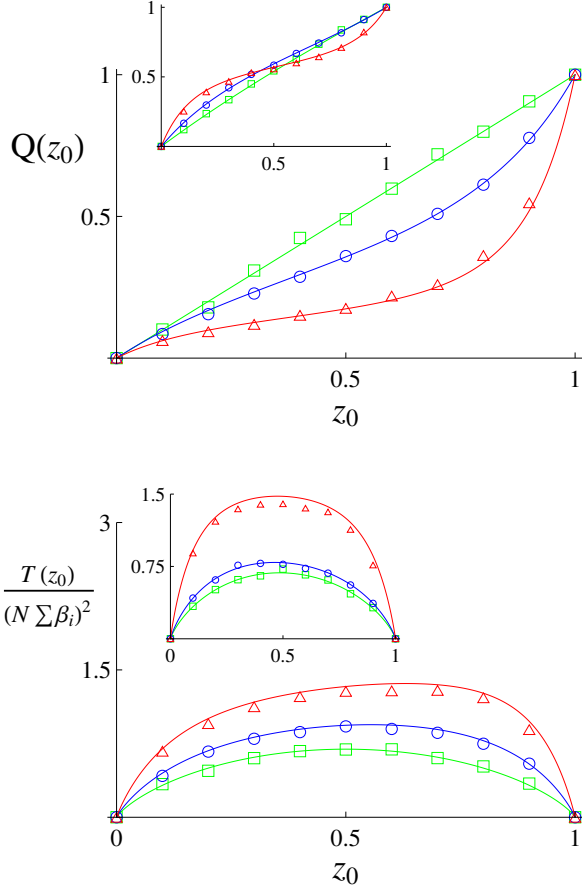


Figure 8: Plots of the probability of fixation,  $Q(z_0)$ , and mean time to fixation,  $T(z_0)$ , as a function of the projected initial conditions, for systems featuring symmetric migration balance. Main plots feature  $\mathcal{D} = 2$ ,  $\beta = (1, 1)$ ,  $\alpha = (1, -1)$  and  $m$  given by Eq. (47) with  $\theta = 0.8$  for increasing values of  $s$ . Continuous lines are obtained from the reduced one-dimensional model, while symbols are obtained from stochastic simulation with  $N = 300$ . Plots in green/squares correspond to  $s = 1.66 \times 10^{-3}$  with  $Q(z_0)$  and  $T(z_0)$  obtained from a first order solution to Eq. (4) and Eq. (6) (see Eq. (44)). The remaining plots are calculated from second order solutions to Eq. (4) and Eq. (6) (see Eq. (43)) with  $s = 5.7 \times 10^{-2}$  for the blue/circle plots and  $s = 0.11$  for the red/triangle plots. Inset plots meanwhile, are obtained from parameters  $\mathcal{D} = 5$ ,  $\beta = (1, 1, 1, 1, 2)$ ,  $\alpha = (1, 1, 2, 2, -1, -1)$  and an  $m$  matrix with an unspecified structure.

system towards fixation at  $\mathbf{x} = (0, 0)$ . The reduced model captures the behaviour extremely well, as observed in figure 8, top panel. The mean time to fixation meanwhile begins to increase, diverging as the deterministic fixed point holds the system in its vicinity for longer and longer. For these very large, arguably unphysical values of  $s$ , the reduced system begins to over-predict the rapidly increasing time to fixation, as seen in figure 8. This can also be seen as a consequence of  $s$  becoming larger than  $|\Re(\lambda^{(2)})|$ .

In this section we have focused on a very restricted set of parameters to illustrate the effect of migration-selection balance. While such stable polymorphic equilibria clearly exist for a host of other parameters, including multiple islands of differing sizes and various selection pressures, the parameter range in which they exist becomes increasingly small relative to the full parameter space as  $\mathcal{D}$  increases. In addition, while the deterministic analysis of such systems becomes progressively more complex, the reduced system continues to provide a good approximation of the fixation probability and fixation time, as demonstrated in figure 8, inset. Finally, we would like to emphasise that we have here applied the reduction method to an extreme and very particular set of parameters, essentially testing the method to breaking point. This is done to demonstrate the quality of the approximation for large values of  $s/N$ .

#### 4.4. Hub

Having discussed the general predictions of the reduced model in both the neutral case and that in which selection is present, we now proceed to apply the results to a specific metapopulation topology, that of the hub or spoke (see figure 9). Our reasons for choosing such a system are twofold. Firstly the system possess symmetries which make it particularly suitable to an analytic treatment (though we stress that our method can also be used for more general systems). Secondly, such a structure allows us to investigate the behaviour of the model systematically as the number of demes increases.

Let us now consider the details of the system. We define the hub topology as one featuring a main deme which is connected to  $\mathcal{D} - 1$  satellite demes. The satellite demes themselves are entirely unconnected to one another. Migration probabilities along the connections are chosen so as to limit the parameter space but still allow for non-trivial behaviour.

Recall the definition of the migration matrix  $m$  in Section 2.2; previously we stated that the columns of  $m$  were normalised such that the probability the offspring from a reproduction event would not migrate was equal to 1 minus the total probability it would migrate,  $m_{jj} = 1 - \sum_{i \neq j} m_{ij}$ . In this case however, since we have a more restricted geometry, we can instead parametrise the migration probabilities by the probability that the offspring *remains* in the same deme as its parent. Defining  $\omega_1$  as the probability that an offspring produced in the central deme does not migrate and  $\omega_2$  the probability that an offspring from a satellite deme does not migrate, the normalised migration matrix for  $\mathcal{D}$  demes is

$$m = \begin{pmatrix} \omega_1 & 1 - \omega_2 & 1 - \omega_2 & \dots & 1 - \omega_2 \\ \frac{1 - \omega_1}{(\mathcal{D} - 1)} & \omega_2 & 0 & \dots & 0 \\ \frac{1 - \omega_1}{(\mathcal{D} - 1)} & 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1 - \omega_1}{(\mathcal{D} - 1)} & 0 & 0 & 0 & \omega_2 \end{pmatrix}. \quad (48)$$



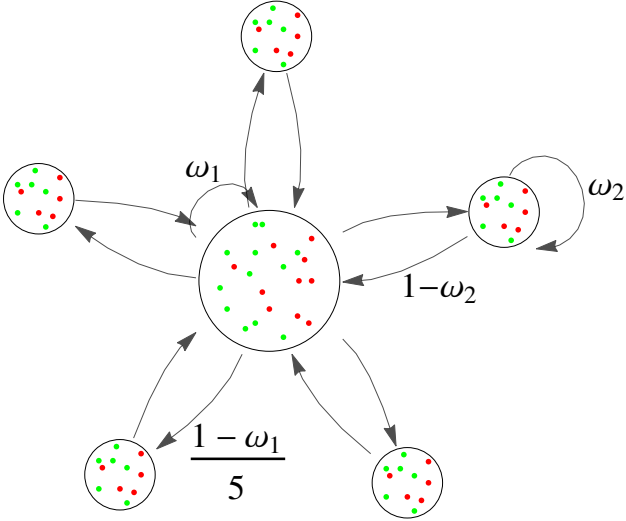


Figure 9: Metapopulation model possessing a hub structure with  $\mathcal{D} = 6$ . The central deme has a population of  $\beta_1 N$  while surrounding demes have a population of  $\beta_2 N$ . Migration probabilities (conditional on origin island being first selected) in this case can be parametrised by the probability of remaining on a particular deme. The probability of remaining on the central deme is  $\omega_1$ , while the probability of migrating is dispersed equally over the satellite demes. The probability of remaining on a satellite deme is  $\omega_2$ , with probability  $1 - \omega_2$  of migrating to the central deme.

Further we take the central deme to have a population of  $\beta_1 N$ , the satellite demes to have populations  $\beta_2 N$ , and the birth rate in each deme be proportional to the island size, so that  $f_j = \beta_j / \sum_{i=1}^{\mathcal{D}} \beta_i$ .

We now apply the theory developed in Section 3. We begin by constructing the matrix  $H$  from the migration matrix  $m$ , and island sizes  $\beta$ . Before proceeding further, we calculate the eigenvalues of  $H$ , as it is these that define the parameter range over which we would expect the approximation to work (see section 3). For convenience we introduce the quantities

$$\gamma_1 = \beta_1^2(1 - \omega_1), \quad (49)$$

$$\gamma_2 = (\mathcal{D} - 1)\beta_2^2(1 - \omega_2), \quad (50)$$

$$\gamma_3 = \gamma_1 + (\mathcal{D} - 1)\gamma_2, \quad (51)$$

$$\gamma_4 = (\mathcal{D} - 1)\beta_1\beta_2[\beta_1 + (\mathcal{D} - 1)\beta_2]. \quad (52)$$

The first two eigenvalues are then given by

$$\lambda^{(1)} = 0, \quad \lambda^{(2)} = -\frac{\gamma_3}{\gamma_4}, \quad (53)$$

and for the remaining eigenvalues we find

$$\lambda^{(j)} = -\frac{\gamma_1}{\gamma_4}, \quad j \geq 3. \quad (54)$$

By considering these eigenvalues we are already alerted to parameter regimes in which the reduced system could potentially give poor agreement with the full system. For instance, as we increase the number of demes in the system,

$\mathcal{D}$ , we find  $\lambda^{(2)}$  tends to a finite quantity,  $-(\omega_2 - 1)/\beta_1$ . However, the remaining non-zero eigenvalues tend to zero with increasing deme number. One therefore must be cautious when applying the approximation technique to a hub system with a large number of satellite demes, as we expect that the approximation will break down if the magnitude of these eigenvalues approaches  $N^{-1/2}$ .

To obtain the reduced model in the neutral case, we need only calculate  $\mathbf{u}^{(1)}$  (see Eq. (33)). In the case where we look at second order effects in  $s$ , we must also calculate the remaining left- and right-eigenvectors. Since the system contains degenerate eigenvalues (Eq. (54)) — this is frequently the case in such highly symmetric systems — the corresponding eigenvectors will not automatically be orthogonal. An orthogonal set must be constructed by taking linear combinations of these vectors, so that the orthonormality condition, Eq. (26), holds. The left-eigenvectors can be expressed as

$$\mathbf{u}^{(1)} = \frac{1}{\gamma_3} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_2 \end{pmatrix}, \quad \mathbf{u}^{(2)} = \begin{pmatrix} -(\mathcal{D} - 1) \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad (55)$$

and

$$u_i^{(j)} = \delta_{ij} - \frac{1}{j-2} \sum_{l=2}^{j-1} \delta_{li}, \quad j \geq 3. \quad (56)$$

The right eigenvectors meanwhile are given by

$$\mathbf{v}^{(1)} = \mathbf{1}, \quad \mathbf{v}^{(2)} = \frac{1}{(\mathcal{D} - 1)\gamma_3} \begin{pmatrix} -(\mathcal{D} - 1)\gamma_2 \\ \gamma_1 \\ \vdots \\ \gamma_1 \end{pmatrix}, \quad (57)$$

and

$$v_i^{(j)} = \frac{j-2}{j-1} \left( \delta_{ij} - \frac{1}{j-2} \sum_{l=2}^{j-1} \delta_{li} \right), \quad j \geq 3. \quad (58)$$

With these quantities in hand we can calculate the fixation probability and fixation time as defined in Section 4.1 and Section 4.2.

#### 4.4.1. Hub: $s=0$

In order to see how our reduced hub model compares against the full system with increasing  $\mathcal{D}$ , we can look at how the fixation probability  $Q(z_0)$  and the fixation time  $T(z_0)$ , normalised by the total system population,  $N_{\text{Tot}}$ , changes with the initial condition  $z_0$ , while other parameters are kept fixed. The results are plotted in figure 10, and we see that as  $\mathcal{D}$  increases the approximation continues to provide good agreement with the exact Gillespie simulation of Eq. (A.1), with transition rates given by Eq. (9) and Eq. (10).

As stated in section 4.1, the probability of fixation is only dependent on network structure through the projected initial condition,  $z_0$ . Since  $z_0$  is held constant in

this case, the probability of fixation does not change as the network structure is altered. We note that the behaviour of the fixation time as a function of the number of demes is non-trivial however. The results in figure 10 may be compared with the results for a single island of the same size as the total hub population; increasing the size of the island would give  $r_N = 1$  regardless of size. Here we see  $r_N$  starts at a significantly higher value and decreases as the deme number (and hence the total population size) is increased.

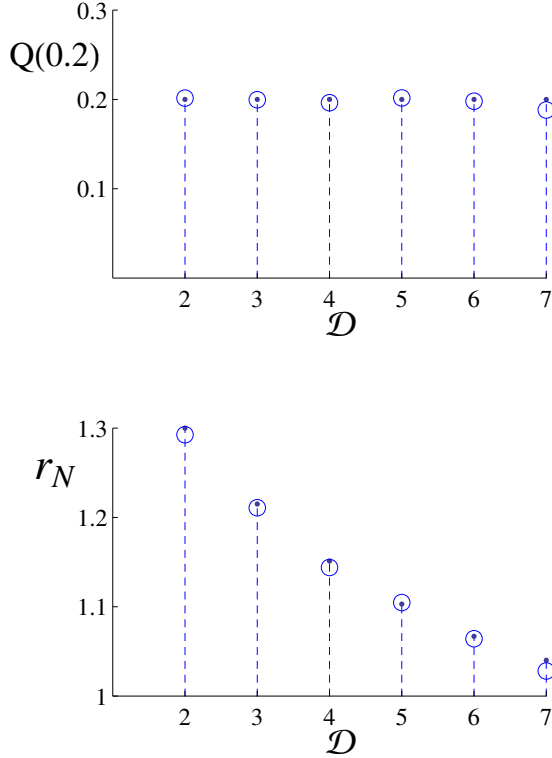


Figure 10: Top panel: probability of fixation,  $Q(z_0)$ , with fixed initial condition  $z_0 = 0.2$ , for the neutral hub model ( $s = 0$ ) plotted as a function of the number of demes,  $\mathcal{D}$ . Bottom panel: parameter  $r_N$  as a function of  $\mathcal{D}$  for the neutral hub model. Discrete analytic results are plotted as a dashed blue columns, and calculated from Eq. (38) and Eq. (39), and Eq. (34) calculated using Eq. (55). Simulation results, plotted as circles, are the mean results from 6000 runs. Parameters used in this example are  $\omega_1 = 0.5$ ,  $\omega_2 = 15/16$ ,  $\beta_1 = 6$ ,  $\beta_2 = 1$ , and finally  $N = 300$ .

#### 4.4.2. Hub with selection

Let us now incorporate selection into the general hub model described in figure 9. Suppose that the selection strength in the central deme, 1, is moderated by  $\alpha_1$  while the selection strength in the satellite demes is moderated by  $\alpha_2$ . The  $\bar{A}(z)$  term for the system is given by Eq. (29) and  $\bar{B}(z)$  by Eq. (33). The parameters  $a_1$ ,  $a_2$  and  $a_3$  are

given by Eqs. (30), (31) and (32) which can now be easily calculated since we have the left- and right-eigenvectors of  $H$  (equations Eq. (55), Eq. (56) and Eq. (57), Eq. (58) respectively). Their exact forms are too lengthy to be reproduced here, but are obtained by direct substitution. The results can then be tested against exact Gillespie simulations of the stochastic system defined by Eq. (18). As an example, let us compare two systems.

In the first system we fix the number of demes to two with the first deme being defined as the central deme with population  $\beta_1 N$  and the second as the satellite deme with population  $\beta_2 N$ . In deme one, the  $A$  alleles experience a selective pressure  $s\alpha_1$ , while in the deme two, the satellite deme, the alleles experience a selective pressure  $s\alpha_2$ .

In the second system, we again have a central deme with a population of  $\beta_1 N$ , but we now fix the total population in each satellite deme to  $N$  and vary the total deme number  $\mathcal{D}$ . Again the fitness in the central deme is equal to  $s\alpha_1$  and the fitness in each satellite deme is equal to  $s\alpha_2$ . We can then say that in both systems, the number of individuals in the selective environments  $s\alpha_1$  and  $s\alpha_2$  are equal if  $\beta_2 = (\mathcal{D} - 1)$ .

Naïvely then, one might expect the systems to behave similarly for such metrics as fixation probability and fixation time, as  $\beta_2$  and  $\mathcal{D} - 1$  respectively increase in each system. However, the analytical results predict very different behaviour. This is supported by simulation; in figure 11, a particular set of parameters is fixed and the size of the populations (moderated by  $\beta_2$  in the first case and discretely by  $\mathcal{D}$  in the second case) is increased. The probability of  $A$  fixating decreases much more rapidly with an increasing number of satellite demes than the two deme system with increasing size of the second island. The time to fixation meanwhile increases more rapidly with the increasing number of satellite demes than that with two islands.

## 5. Conclusions

Migration, along with mutation, selection and drift is one of the fundamental processes occurring in genetics, and island models were introduced very early on in the mathematical formulation of population genetics [44]. Yet theoretical studies of their general features are comparatively rare, and even the literature on specific models, while quite extensive, features no common approach. Given the wide variety of methodologies and notations employed, it is particularly difficult to get an overview of theoretical work on island models. A major reason for this must be the perceived difficulty of formulating a stochastic version of the theory — required if the population size is not assumed to be infinite — and even more, the intractability of the resulting equations.

In this paper we have tried to address both of these problems. The first, that of formulating a stochastic version of the model, is the least difficult. We contend that the most systematic and straightforward way of proceeding is to first formulate the model at the ‘microscopic’ level, that



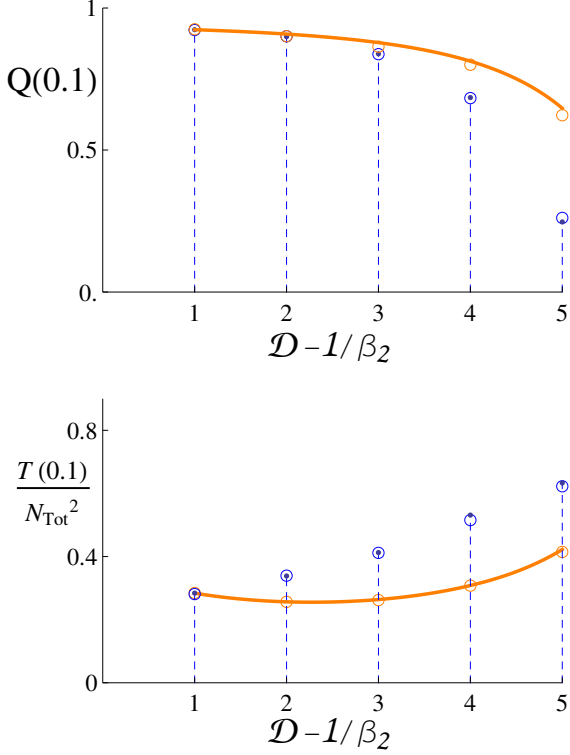


Figure 11: Plots of fixation probability,  $Q(z_0)$  and fixation time,  $T(z_0)$  for two different models at  $z_0 = 0.1$ . The orange lines are obtained from the reduced model of a two-deme system in which  $\beta_2$ , the relative size of the second deme, is increased. The results from simulation of the system are shown as orange circles. The discrete values indicated by the blue dashed lines are obtained from a hub model in which the number of  $(D-1)$  satellite demes is increased. Simulation results are plotted as blue circles. Simulation results are the mean of 2000 runs. Parameters used are  $s = 0.03$ ,  $\alpha_1 = 1$ ,  $\alpha_2 = -1$ ,  $\beta_1 = 3$ ,  $\beta_2 = 1$ ,  $\omega_1 = 0.625$ ,  $\omega_2 = 0.9375$ , and  $N_{\text{Tot}} = 300$ .

is, in terms of individuals which undergo the fundamental processes of birth, death, migration between islands, etc.. This can be achieved through the definition of states of the system (simply the numbers of individuals of different types on each island) and transition probabilities per unit time between these states. The latter are made up of combinatoric factors corresponding to sampling of specific types of individuals in the system, together with the rates at which the fundamental processes such as birth, death and migration occur. These transition probabilities per unit time, together with the usual assumption that the processes are Markovian, lead to a master equation (a Markov chain in continuous time) which governs the dynamics of the system.

This is a systematic procedure which results in an equation which completely specifies the stochastic dynamics, once initial conditions are given. However, the master equation is completely intractable in all situations of in-

terest, and so starting from the earliest studies [15] a diffusion approximation has been made. Here the states are now the fraction of individuals of different types, which for large island sizes can be taken to be continuous variables to a good approximation. Within this approximation the master equation becomes a FPE which is equivalent to a set of SDEs. This is the starting point for many authors in their studies of island models. However, we stress again that beginning from a master equation is both more natural and easier to interpret. It also avoids technical issues, such as trying to decide whether or not the process is of the Itô or Stratonovich type [17]. This procedure was described in [5] for the neutral case and islands all having the same number of individuals. Here we have extended it to include the effects of selection and differing island size.

The equations resulting from this procedure are, however, still formidably complicated. The FPEs are multidimensional partial differential equations and the set of SDEs are coupled, nonlinear and have multiplicative noise. The second, and more substantive, aim of this paper has been to show that the application of a procedure which we recently devised [7], allows the FPE to be reduced to a much simpler FPE, or equivalently allows the set of SDEs to be reduced to a single SDE. The resulting equations are then amenable to analysis. The technique is simple to understand, being based on the elimination of modes in the system which decay quickly, and gives rise to explicit formulae for the parameters in the final, simplified equations. These parameters can be straightforwardly calculated from the network structure of the islands.

We have applied this method to systems defined by a range of different networks, comparing the results we obtain to a direct simulation of the IBM. The method gives excellent results for most networks and parameter values. Where it does not work so well, there are reasons why we would expect this. For instance, we require the magnitude of the real part of the non-zero eigenvalues of  $H$  to be greater than both  $s$  and  $N^{-1/2}$ , so that the separation of timescales is sufficient to apply the approximation. Furthermore, we assume that parameters modelling similar effects in the model are not of completely different orders. Therefore, no island is assumed to be an order or two of magnitude bigger than other islands (all  $\beta_i$  are of order one). Even if these conditions are violated, the approximation may work reasonably well. We expect that the elimination of fast variables will still be possible in such cases, so a different calibration could describe some of these situations. For example, the diagonal elements of  $H$  could be made to scale with  $N^{-1/2}$ , which would result in a different set of formulae. In [1], the diagonal elements were chosen to scale like  $N^{-1}$ , which accounts in part for the differences between the results given in that paper and those given here.

We have chosen to work only to order  $s^2$ , although the technique is capable of being generalised to higher orders in  $s$ . Given the typical size of selection strengths, working to this order is entirely reasonable, and indeed most authors

only keep terms of order  $s$ . We should emphasise that although we frequently compare  $s$  numerically to  $N^{-1/2}$  or  $N^{-1}$ , we do not, unlike some authors, set  $s$  equal to  $N^{-1/2}$  or  $N^{-1}$ ;  $s$  and  $N$  are two independent parameters. One of the reasons we go to order  $s^2$  is to show that the analysis of the stochastic aspects of migration-selection balance could be considerably extended using our results.

The model we present here is slightly different from the previous work we have addressed which concerns migration-selection balance. However, we have shown that our method, being more general, can be used to investigate a broader range of parameters than previously attempted, including non-symmetric migration, arbitrary deme topology and an arbitrary range of selective pressures for each deme. Further, we note that the mechanism that allows our approximation to work so successfully is the dominance of the large and linear effect of migration (in our model embodied by the matrix  $H$ ) over smaller non-linear and stochastic terms.

Conceivably then, in addition to the points we have mentioned, there may be other ways of extending and generalising our treatment. For instance, we have imposed fixed population size for each island. This is traditional in the context of population genetics, and the Moran process has this assumption at its heart. However it should be possible to relax this assumption. Other processes, such as mutation, could also be included. We are in the process of investigating some of the questions, and hope to report on them in the future.

The formulation and subsequent analysis of stochastic effects or of migration in population genetics is often perceived as being difficult, and systems where both are important doubly so. We believe this need not be the case. The formulation can be made systematic and intuitively appealing by starting from an IBM and invoking the diffusion approximation. The analysis of the resulting equations is possible through the elimination of the fast variables, leading to much simpler equations which are amenable to analysis. We hope that this methodology expounded here will be taken up by other researchers and will lead to the analysis of more complex and realistic models.

## Acknowledgement

We thank Diana Garcia López for useful discussions. G.W.A.C. thanks the Faculty of Engineering and Physical Sciences, University of Manchester for funding through a Dean's Scholarship.

## Appendix A. The transition rates and the derivation of the Fokker-Planck equation

Here we describe the mathematical machinery employed in moving from a description of the system in terms of the probability transition rates (e.g. Eqs. (9) and (10)), to a

description in terms of the Fokker-Planck equation or the equivalent stochastic differential equations.

The state of the system at any moment in time is given by the vector  $\mathbf{n} = (n_1, \dots, n_D)$ , where  $n_i$  specifies the number of individuals carrying allele  $A$  on island  $i$ . The transition rates  $T(\mathbf{n}|\mathbf{n}')$ , are measures of the probability per unit time that the system moves from state  $\mathbf{n}'$  to  $\mathbf{n}$ . The transition rates for the neutral migration model are given by Eqs. (9) and (10), where they are derived from a consideration of some standard combinatorics. Here we shall discuss the derivation of the model including selection (see Section 2.4), which reduces to the neutral model as the selection strength  $s$  tends to zero.

Suppose that now, rather than individuals carrying allele  $A$  or  $B$  being just as likely to reproduce, that one has a fitness advantage over the other. In an analogous manner to the one deme case of Section 2.3, we introduce weighting vectors  $\mathbf{w}_A$  and  $\mathbf{w}_B$  which describe the relative likelihood of type  $A$  or  $B$  reproducing on any of the islands. The probability per unit time of a type  $A$  individual on island  $i$  reproducing is then given by the number of  $A$  individuals on island  $i$ ,  $n_i$ , multiplied by the relative fitness of allele  $A$  on island  $i$ ,  $[\mathbf{w}_A]_i$ , normalised by the total fitness of the population of island  $i$ ,  $n_i[\mathbf{w}_A]_i + (N - n_i)[\mathbf{w}_B]_i$ . The progeny may then either remain in its own deme or migrate to another, based on the migration rate matrix  $G$ . The individual it displaces is chosen neutrally, based on local allele frequencies. The probability transition rates are then given by

$$\begin{aligned} T(n_i + 1 | n_i) &= \sum_{j=1}^D \frac{(\beta_i N - n_i)}{\beta_i N - \delta_{ij}} G_{ij} \frac{[\mathbf{w}_A]_j n_j}{[\mathbf{w}_A]_j n_j + [\mathbf{w}_B]_j (\beta_j N - n_j)}, \\ T(n_i - 1 | n_i) &= \sum_{j=1}^D \frac{n_i}{\beta_i N - \delta_{ij}} G_{ij} \frac{[\mathbf{w}_B]_j (\beta_j N - n_j)}{[\mathbf{w}_A]_j n_j + [\mathbf{w}_B]_j (\beta_j N - n_j)}, \end{aligned}$$

using analogous arguments to those used to obtain the neutral transition rates.

We can further simplify these expressions by setting  $[\mathbf{w}_B]_i = 1$  and  $[\mathbf{w}_A]_i = 1 + s\alpha_i$  for each island. The parameter  $s$  is an indicative selection strength, while the elements of  $\alpha$  will be assumed to be of order 1 and will primarily be used to signify the direction of selection. If  $\alpha_i > 0$  then  $[\mathbf{w}_A]_i > [\mathbf{w}_B]_i$  and allele  $A$  is advantageous on island  $i$ , while if  $\alpha_i < 0$ , allele  $A$  will be deleterious on that island. Finally, if we assume that the selection strength  $s$  is small, we can express the above transition rates as a Taylor series in  $s$ . Suppressing the dependence of  $T(\mathbf{n}|\mathbf{n}')$  on states that do not vary in a particular transition, we

obtain

$$\begin{aligned}
T(n_i + 1|n_i) &= \sum_{j=1}^{\mathcal{D}} \frac{(\beta_i N - n_i)}{\beta_i N - \delta_{ij}} G_{ij} \times \\
&\quad \left( \frac{n_j}{\beta_j N} + s \alpha_j \frac{n_j(\beta_j N - n_j)}{(\beta_j N)^2} \right. \\
&\quad \left. - s^2 \alpha_j^2 \frac{n_j^2(\beta_j N - n_j)}{(\beta_j N)^3} + \mathcal{O}(s^3) \right), \\
T(n_i - 1|n_i) &= \sum_{j=1}^{\mathcal{D}} \frac{n_i}{\beta_i N - \delta_{ij}} G_{ij} \times \\
&\quad \left( 1 - \frac{n_j}{\beta_j N} - s \alpha_j \frac{n_j(\beta_j N - n_j)}{(\beta_j N)^2} \right. \\
&\quad \left. + s^2 \alpha_j^2 \frac{n_j^2(\beta_j N - n_j)}{(\beta_j N)^3} + \mathcal{O}(s^3) \right).
\end{aligned}$$

The dynamics of the system can be described by a master equation, as explained Section 2.1 in the case of a single island without selection. It is a simple generalisation of Eq. (1) with now the state being specified by the vector  $\mathbf{n}$ :

$$\begin{aligned}
\frac{dp(\mathbf{n}, t)}{dt} &= \sum_{i=1}^{\mathcal{D}} [T(n_i|n_i - 1)p(n_i - 1, t) \\
&\quad - T(n_i + 1|n_i)p(n_i, t)] \\
&+ \sum_{i=1}^{\mathcal{D}} [T(n_i|n_i + 1)p(n_i + 1, t) \\
&\quad - T(n_i - 1|n_i)p(n_i, t)]. \quad (\text{A.1})
\end{aligned}$$

It is to this master equation, with the transition rates given above, that we wish to apply the Kramers-Moyal expansion to obtain the Fokker-Planck equation.

The dynamics can be seen to be that of a one-step process; any one transition can only move the system from an initial state  $\mathbf{n}' = (n_1, \dots, n_i, \dots, n_{\mathcal{D}})$  to the adjacent states  $\mathbf{n} = (n_1, \dots, n_i \pm 1, \dots, n_{\mathcal{D}})$ . We can exploit this fact notationally; introducing new state variables  $\mathbf{x}$  such that  $x_i = n_i/\beta_i N$ , we can write  $f_i^+(x_i)$  and  $f_i^-(x_i)$  as shorthand for the transition rates (in terms of the new variables) for moving up to state  $x_i + 1/\beta_i N$  or down in

state  $x_i - 1/\beta_i N$  from initial state  $\mathbf{x}'$ . This gives

$$\begin{aligned}
f_i^+(x_i) &= \frac{G_{ii}(1 - x_i)}{1 - (\beta_i N)^{-1}} \times \\
&\quad [x_i + s \alpha_i x_i(1 - x_i) - s^2 \alpha_i^2 x_i^2(1 - x_i)] \\
&\quad + (1 - x_i) \sum_{j \neq i}^{\mathcal{D}} G_{ij} [x_j + s \alpha_j x_j(1 - x_j) \\
&\quad \quad - s^2 \alpha_j^2 x_j^2(1 - x_j)] + \mathcal{O}(s^3), \\
f_i^-(x_i) &= \frac{G_{ii}x_i}{1 - (\beta_i N)^{-1}} \times \\
&\quad [(1 - x_i) - s \alpha_i x_i(1 - x_i) + s^2 \alpha_i^2 x_i^2(1 - x_i)] \\
&\quad + x_i \sum_{j \neq i}^{\mathcal{D}} G_{ij} [(1 - x_j) - s \alpha_j x_j(1 - x_j) \\
&\quad \quad + s^2 \alpha_j^2 x_j^2(1 - x_j)] + \mathcal{O}(s^3). \quad (\text{A.2})
\end{aligned}$$

For now let us leave the specific form of these transition rate functions alone, pausing only to note that the typical deme size,  $N$ , now only appears in the first term of  $f_i^+(x_i)$  and  $f_i^-(x_i)$ .

We now re-express the master equation in terms of the transition rates  $f_i^+(x_i)$  and  $f_i^-(x_i)$ :

$$\begin{aligned}
\frac{dp}{dt} &= \sum_{i=1}^{\mathcal{D}} \left[ f_i^+(x_i - \frac{1}{\beta_i N}) p(x_i - \frac{1}{\beta_i N}, t) \right. \\
&\quad \left. - f_i^+(x_i) p(x_i, t) \right] \\
&+ \sum_{i=1}^{\mathcal{D}} \left[ f_i^-(x_i + \frac{1}{\beta_i N}) p(x_i + \frac{1}{\beta_i N}, t) \right. \\
&\quad \left. - f_i^-(x_i) p(x_i, t) \right]. \quad (\text{A.3})
\end{aligned}$$

Assuming the typical deme population  $N$  to be large, we can carry out a Taylor expansion in  $N^{-1}$ ; this is in effect the Kramers-Moyal expansion [17, 35]. The right-hand side of the master equation (A.3) becomes

$$\begin{aligned}
&- \sum_{i=1}^{\mathcal{D}} \left\{ \left( \frac{1}{\beta_i N} \right) \frac{\partial}{\partial x_i} [f_i^+(x_i) p(x_i, t)] \right\} \\
&+ \frac{1}{2!} \sum_{i=1}^{\mathcal{D}} \left\{ \left( \frac{1}{\beta_i N} \right)^2 \frac{\partial^2}{\partial x_i^2} [f_i^+(x_i) p(x_i, t)] \right\} \\
&+ \sum_{i=1}^{\mathcal{D}} \left\{ \left( \frac{1}{\beta_i N} \right) \frac{\partial}{\partial x_i} [f_i^-(x_i) p(x_i, t)] \right\} \\
&+ \frac{1}{2!} \sum_{i=1}^{\mathcal{D}} \left\{ \left( \frac{1}{\beta_i N} \right)^2 \frac{\partial^2}{\partial x_i^2} [f_i^-(x_i) p(x_i, t)] \right\},
\end{aligned}$$

plus terms in  $N^{-3}$  and higher.

We now return to the terms in  $f_i^+$  and  $f_i^-$  which involve  $N$ . They are identical to lowest order in  $s$ , and equal

$$\frac{G_{ii}(1 - x_i)x_i}{1 - (\beta_i N)^{-1}}. \quad (\text{A.4})$$

Now in the master equation  $f_i^+$  and  $f_i^-$  appear with different signs in the terms involving the first derivative, and so they cancel. Although their contributions add in the terms involving the second derivative, if we expand the expression (A.4) in powers of  $N^{-1}$  we see that these give  $\mathcal{O}(N^{-3})$  contributions in the Kramers-Moyal expansion, which we are discarding. By the same argument, the terms in  $f_i^+$  and  $f_i^-$  which involve  $N$  and powers of  $s$  will also give  $\mathcal{O}(N^{-3})$  contributions when multiplying the second derivative, and so can also be discarded. Finally, when these  $s$ -dependent terms multiply the first derivative, they will give contributions  $s/N^2$  and  $s^2/N^2$ , but we will not include such terms in the diffusion matrix  $B$  (see below), and so we do not include them in this context either. So, in summary, the  $N$  dependence which appears in  $f_i^+$  and  $f_i^-$  in Eq. (A.2) may be omitted to the order we are working, and the only  $N$  dependence is that shown explicitly in the Kramers-Moyal expansion of the right-hand side of the master equation.

We now define

$$\begin{aligned} A_i(\mathbf{x}) &= \frac{1}{\beta_i} [f_i^+(\mathbf{x}) - f_i^-(\mathbf{x})], \\ B_{ii}(\mathbf{x}) &= \frac{1}{\beta_i^2} [f_i^+(\mathbf{x}) + f_i^-(\mathbf{x})]. \end{aligned} \quad (\text{A.5})$$

With these definitions the expansion of the master equation in inverse powers of  $N$  takes the form

$$\begin{aligned} \frac{\partial p(\mathbf{x}, t)}{\partial t} = & - \frac{1}{N} \sum_{i=1}^{\mathcal{D}} \frac{\partial}{\partial x_i} [A_i(\mathbf{x}) p(\mathbf{x}, t)] \\ & + \frac{1}{2N^2} \sum_{i=1}^{\mathcal{D}} \frac{\partial^2}{\partial x_i^2} [B_{ii}(\mathbf{x}) p(\mathbf{x}, t)]. \end{aligned} \quad (\text{A.6})$$

Substituting the explicit forms for  $f_i^\pm$  given by Eq. (A.2) into Eq. (A.5) gives the elements of the vector  $\mathbf{A}(\mathbf{x})$  as

$$\begin{aligned} A_i(\mathbf{x}) = & \frac{1}{\beta_i} \left\{ \sum_{j \neq i}^{\mathcal{D}} G_{ij} (x_j - x_i) \right. \\ & + s \sum_{j=1}^{\mathcal{D}} G_{ij} \alpha_j x_j (1 - x_j) \\ & \left. - s^2 \sum_{j=1}^{\mathcal{D}} G_{ij} \alpha_j^2 x_j^2 (1 - x_j) \right\} + \mathcal{O}(s^3), \end{aligned}$$

and a diagonal diffusion matrix with elements given by

$$\begin{aligned} B_{ii}(\mathbf{x}) = & \frac{1}{\beta_i^2} \left\{ x_i \sum_{j=1}^{\mathcal{D}} G_{ij} + \sum_{j=1}^{\mathcal{D}} G_{ij} x_j \right. \\ & \left. - 2x_i \sum_{j=1}^{\mathcal{D}} G_{ij} x_j \right\} + \mathcal{O}(s). \end{aligned} \quad (\text{A.7})$$

The truncation of the series in  $s$ , should be chosen to be consistent with the truncation in the expansion in terms

of  $N$ . This will clearly depend on the assumed size of  $s$ . If one sets  $s = 0$ , the above model reduces to that stated for the neutral case, Eq. (12) and Eq. (13). Further, it can be shown that the Fokker-Planck equation is equivalent [17, 35] to the SDE

$$\frac{dx_i}{d\tau} = A_i(\mathbf{x}) + \frac{1}{\sqrt{N}} \eta_i(\tau), \quad (\text{A.8})$$

defined in the sense of Itô [40]. Here  $\tau = t/N$ ,  $\boldsymbol{\eta}(\tau)$  is a Gaussian white noise term such that  $\langle \eta_i(\tau) \rangle = 0$  and  $\langle \eta_i(\tau) \eta_j(\tau') \rangle = B_{ij}(\mathbf{x}) \delta(\tau - \tau')$ , although for the present model  $B_{ij}(\mathbf{x}) = 0$  if  $i \neq j$ .

## Appendix B. Solution to the equation for the probability of fixation

The probability of fixation in the reduced system,  $Q(z_0)$ , is found from the backward equation corresponding to the FPE (36), in exactly the same way that the equation for the probability of fixation in the single island case (4), is found from the backward equation corresponding to Eq. (2). Therefore the equation reads

$$\frac{\bar{A}(z_0)}{N} \frac{dQ}{dz_0} + \frac{\bar{B}(z_0)}{2N^2} \frac{d^2Q}{dz_0^2} = 0, \quad (\text{B.1})$$

where  $z_0$  is the initial starting point on the centre manifold (or slow subspace). The boundary conditions are as for the single island case, that is,  $Q(0) = 0$  and  $Q(1) = 1$ . In this appendix we discuss the analytic solution of Eq. (B.1) when  $\bar{A}(z_0)$  and  $\bar{B}(z_0)$  are given by Eq. (29) and Eq. (33).

The result for the neutral case and to linear order in  $s$  have the same form as in the one-island case, and are well known [13]. When  $s = 0$ ,  $\bar{A}(z_0) = 0$ , and so the solution of Eq. (B.1) subject to the boundary conditions is simply  $Q(z_0) = z_0$ . At linear order in  $s$ ,  $\bar{A}(z) = sa_1 z(1 - z)$ , and a straightforward integration of Eq. (B.1) gives Eq. (17), albeit with extra factors of  $a_1$  and  $b_1$  and with  $x_0$  replaced by  $z_0$  (see Eqs. (43) and (44)).

To second order in  $s$ ,  $\bar{A}(z)$  may be written in the form (40), while  $\bar{B}(z)$  is still given by Eq. (33). The equation for the probability of fixation (B.1) now takes the form

$$\frac{s}{N} z_0(1 - z_0)(k_1 - sk_2 z_0) \frac{dQ}{dz_0} + \frac{1}{N^2} b_1 z_0(1 - z_0) \frac{d^2Q}{dz_0^2} = 0.$$

Integrating with respect to  $z_0$  we arrive at the equation

$$\frac{dQ}{dz_0} = c_1 \exp \left[ -\frac{Ns}{b_1} (k_1 z_0 - \frac{sk_2}{2} z_0^2) \right],$$

where  $c_1$  is a constant of integration yet to be determined and where we note from Eq. (34) that  $b_1 > 0$ .

If  $k_2 = 0$ , the calculation is identical to that carried out to first order in  $s$ , Eq. (44), but with  $a_1$  replaced by  $k_1$ . If  $k_2 \neq 0$ , we may complete the square in the exponent to find

$$\frac{dQ}{dz_0} = c_1 \exp \left[ -\frac{Nk_1^2}{2b_1k_2} \right] \exp \left[ \frac{N}{2b_1k_2} (sk_2 z_0 - k_1)^2 \right].$$

We now change variables from  $z_0$  to  $l$ , where

$$l = \sqrt{\frac{N}{2b_1|k_2|}}(sk_2z_0 - k_1), \quad (\text{B.2})$$

to obtain

$$\frac{dQ}{dl} = \begin{cases} -c_2 \exp(-l^2), & \text{if } k_2 < 0 \\ c_2 \exp(l^2), & \text{if } k_2 > 0, \end{cases} \quad (\text{B.3})$$

where

$$c_2 = \frac{c_1}{s} \sqrt{\frac{2b_1}{|k_2|N}} \exp\left\{-\frac{Nk_1^2}{2b_1k_2}\right\}, \quad (\text{B.4})$$

is another constant.

The integrals over the exponentials in Eq. (B.3) can be carried out in terms of functions related to the error function, namely the complementary error function [29]

$$\text{erfc}y = 1 - \text{erf}y = 1 - \frac{2}{\sqrt{\pi}} \int_0^y e^{-l^2} dl, \quad (\text{B.5})$$

and the imaginary error function [10]

$$\text{erfi}y = \frac{2}{\sqrt{\pi}} \int_0^y e^{l^2} dl. \quad (\text{B.6})$$

Implementing the boundary conditions  $Q(l(z_0 = 0)) = 0$  and  $Q(l(z_0 = 1)) = 1$ , one finds

$$Q(z_0) = \frac{1 - \chi(z_0)}{1 - \chi(1)}, \quad (\text{B.7})$$

where

$$\chi(z_0) = \frac{\text{erfc}(l(z_0))}{\text{erfc}(l(0))}, \quad \text{if } k_2 < 0, \quad (\text{B.8})$$

and

$$\chi(z_0) = \frac{\text{erfi}(l(z_0))}{\text{erfi}(l(0))}, \quad \text{if } k_2 > 0. \quad (\text{B.9})$$

If  $l$  is large, then asymptotic forms can be used to simplify both the complementary error function and the imaginary error function [29, 10]:

$$\text{erfc}(l) = \frac{e^{-l^2}}{\sqrt{\pi}l} \left[1 + \mathcal{O}\left(\frac{1}{l^2}\right)\right], \quad (\text{B.10})$$

and

$$\text{erfi}(l) = \frac{e^{l^2}}{\sqrt{\pi}l} \left[1 + \mathcal{O}\left(\frac{1}{l^2}\right)\right]. \quad (\text{B.11})$$

### Appendix C. Calculation of the mean time to fixation

The mean time to fixation in the reduced system,  $T(z_0)$ , is found from the backward equation corresponding to the FPE (36), in exactly the same way that the equation for mean time to fixation in the single island case

(6), is found from the backward equation corresponding to Eq. (2). Therefore the equation reads

$$\frac{\bar{A}(z_0)}{N} \frac{dT}{dz_0} + \frac{\bar{B}(z_0)}{2N^2} \frac{d^2T}{dz_0^2} = -1, \quad (\text{C.1})$$

where  $z_0$  is the initial starting point on the centre manifold (or slow subspace). The boundary conditions are as for the single island case, that is,  $T(0) = 0$  and  $T(1) = 0$ . In this appendix we discuss the analytic solution of Eq. (C.1) when  $\bar{A}(z_0)$  and  $\bar{B}(z_0)$  are given by Eq. (29) and Eq. (33).

The result for the neutral case is well known [13]. Setting  $s = 0$  in Eq. (29) gives  $\bar{A}(z_0) = 0$ , and direct integration of Eq. (C.1) gives Eq. (8), albeit divided by a factor of  $b_1$  and with  $x_0$  replaced by  $z_0$ . At order  $s$ ,  $\bar{A}(z) = sa_1z(1-z)$ , and so the equation for  $T(z_0)$  becomes

$$\frac{\sigma z_0(1-z_0)}{M} \frac{dT}{dz_0} + \frac{z_0(1-z_0)}{M^2} \frac{d^2T}{dz_0^2} = -1, \quad (\text{C.2})$$

where we have defined new parameters  $M = N/\sqrt{b_1}$  and  $\sigma = a_1s/\sqrt{b_1}$ . The reason for introducing these new parameters, other than on grounds of simplicity, is that Eq. (C.2) is exactly the equation found in the single island case with selection.

To solve it we introduce  $\phi(z_0) = dT/dz_0$ , so that the equation now reads

$$\frac{d\phi}{dz_0} + M\sigma\phi = -\frac{M^2}{z_0(1-z_0)}. \quad (\text{C.3})$$

This equation is difficult to deal with analytically and numerically because of the singularities on the right-hand side at precisely the values of  $z_0$  where we need to impose the boundary conditions. One can avoid this problem by writing  $\phi = \phi_0 + \phi_s$ , and choosing  $\phi_0$  so that the term  $d\phi_0/dz_0$  cancels the right-hand side of Eq. (C.3). This choice means that  $\phi_0$  is simply the  $s = 0$  solution, and the equation for  $\phi_s$  is then

$$\frac{d\phi_s}{dz_0} + M\sigma\phi_s = -M\sigma\phi_0 = M^3\sigma[\ln z_0 - \ln(1-z_0)], \quad (\text{C.4})$$

which on the left-hand side is exactly the same as the equation for  $\phi$ , but with a right-hand side which is less divergent as  $z_0 \rightarrow 0$  or  $z_0 \rightarrow 1$ . Although this right-hand side is still divergent, its integral is not, which is all that we need. If we do require a convergent expression we can repeat the process, and write  $\phi_s = \phi_1 + \phi_2$ , choosing  $\phi_1$  so that the term  $d\phi_1/dz_0$  cancels the right-hand side of Eq. (C.4).

We can now multiply Eq. (C.4) by  $e^{M\sigma z_0}$  to find

$$\frac{d}{dz_0} [e^{M\sigma z_0} \phi_s] = M^3\sigma[\ln z_0 - \ln(1-z_0)] e^{M\sigma z_0}, \quad (\text{C.5})$$

which allows the integration to be straightforwardly carried out. One finds

$$T_s(z_0) = c_1 e^{-M\sigma z_0} + c_2 + M^3\sigma \int_0^{z_0} dy e^{-M\sigma y} \int_0^y dx e^{M\sigma x} [\ln x - \ln(1-x)], \quad (\text{C.6})$$

where  $T_s$  is such that  $dT_s/dz_0 = \phi_s$  and  $c_1$  and  $c_2$  are integration constants. Before imposing the boundary conditions, we can simplify the double integral by differentiating the inner integral and integrating by parts. This gives

$$\begin{aligned} T_s(z_0) = & c_1 e^{-M\sigma z_0} + c_2 \\ & - M^2 e^{-M\sigma z_0} \int_0^{z_0} dx e^{M\sigma x} [\ln x - \ln(1-x)] \\ & + M^2 \int_0^{z_0} dy [\ln y - \ln(1-y)]. \end{aligned} \quad (\text{C.7})$$

The last term in Eq. (C.7) is simply the  $s = 0$  mean time to fixation, and so applying the boundary conditions one obtains the equations (45) and (46) given in the main text.

The calculation of  $T(z_0)$  when  $\bar{A}(z_0)$  is taken to order  $s^2$  can be carried out in a similar way, but the results are more complicated and an integration by parts cannot straightforwardly simplify the double integral down to a single integral. The analogous equation to (C.3) is

$$\frac{d\phi}{dz_0} + M\sigma(1 - s\kappa z_0)\phi = -\frac{M^2}{z_0(1-z_0)}, \quad (\text{C.8})$$

where  $\kappa = k_2/k_1$  and  $\sigma$  is now given by  $\sigma = k_1 s/\sqrt{b_1}$ . This is just as singular as Eq. (C.3), and so we perform the same manoeuvre and write  $\phi = \phi_0 + \phi_s$ , choosing  $\phi_0$  so that the term  $d\phi_0/dz_0$  cancels the right-hand side of Eq. (C.8). The equation for  $\phi_s$  then reads

$$\frac{d\phi_s}{dz_0} + M\sigma(1 - s\kappa z_0)\phi_s = M^3\sigma(1 - s\kappa z_0) [\ln z_0 - \ln(1-z_0)]. \quad (\text{C.9})$$

The right-hand side is now less divergent, and one can proceed as before to multiply this equation by  $e^{M\sigma(z_0 - s\kappa z_0^2/2)}$  and integrate twice. We find

$$\begin{aligned} T(z_0) = & -M^2 [z_0 \ln(z_0) + (1-z_0) \ln(1-z_0)] \\ & + M^3\sigma \int_0^{z_0} dy e^{-M\sigma y(1-s\kappa y/2)} \left\{ \int_0^y dx (1-s\kappa x) \times \right. \\ & \left. e^{M\sigma x(1-s\kappa x/2)} \int_0^y [\ln x - \ln(1-x)] - c_3 \right\}, \end{aligned} \quad (\text{C.10})$$

where the constant  $c_3$  is given by

$$\begin{aligned} c_3 = & \left( \int_0^1 dy e^{-M\sigma y(1-s\kappa y/2)} \right)^{-1} \int_0^1 dy e^{-M\sigma y(1-s\kappa y/2)} \\ & \times \int_0^y dx (1-s\kappa x) e^{M\sigma x(1-s\kappa x/2)} [\ln x - \ln(1-x)]. \end{aligned}$$

## References

- [1] Baxter, G. J., Blythe, R. A., Croft, W., & McKane, A. J. (2006). *Phys. Rev. E* **73**, 046118.
- [2] Baxter, G. J., Blythe, R. A., & McKane, A. J. (2008). *Phys. Rev. Lett.* **101**, 258701.
- [3] Baxter, G. J., Blythe, R. A., & McKane, A. J. (2012). *Phys. Rev. E* **86**, 031142.
- [4] Blanquart, F., Gandon, S., & Nuismer, S. L. (2012). *J. Evol. Biol.* **25**, 1351–1363.
- [5] Blythe, R. A. & McKane, A. J. (2007). *J. Stat. Mech.*, P07018.
- [6] Charlesworth, B. (2009). *Nature Reviews Genetics*, **10**, 195–205.
- [7] Constable, G. W. A. & McKane, A. J. (2014). Fast-mode elimination in stochastic metapopulation models. Pre-print arXiv:1402.1355.
- [8] Crow, J. F. & Kimura, M. (1956). In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, (Neyman, J., ed) pp. 1–22. University of California Press Berkeley.
- [9] Crow, J. F. & Kimura, M. (2009). *An Introduction to Population Genetics Theory*. Caldwell, New Jersey, USA: The Blackburn Press.
- [10] Erdélyi, A., ed (1953). *Higher Transcendental Functions: Vol II*. New York: McGraw-Hill.
- [11] Ethier, S. N. & Nagylaki, T. (1980). *Adv. Appl. Probab.* **12**, 14–49.
- [12] Ewens, W. J. (1969). *Population Genetics*. Birkenhead: Wilmer Brother Limited.
- [13] Ewens, W. J. (2004). *Mathematical Population Genetics: I. Theoretical Introduction*. Berlin: Springer-Verlag. Second edition.
- [14] Eyland, E. A. (1971). *Genetics*, **69**, 399–403.
- [15] Fisher, R. A. (1922). *Proc. Roy. Soc. Edin.* **42**, 321–341.
- [16] Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
- [17] Gardiner, C. W. (2009). *Handbook of Stochastic Methods*. Berlin: Springer, Fourth edition.
- [18] Gavrilits, S. & Gibson, N. (2002). *Popul. Ecol.* **44**, 51–58.
- [19] Gillespie, D. T. (1976). *J. Comput. Phys.* **22**, 403–434.
- [20] Hartl, D. L. & Clark, A. G. (2007). *Principles of Population Genetics*. Sunderland, Mass.: Sinauer Associates Inc. Fourth edition.
- [21] Kimura, M. (1955). *Cold Spring Harbour Symp. Quant. Biol.* **20**, 33–53.
- [22] Kimura, M. (1964). *J. Appl. Probab.* **1**, 177–232.
- [23] Kimura, M. (1994). *Population Genetics, Molecular Evolution and the Neutral Theory*. Chicago: Chicago University Press.
- [24] Kimura, M. & Weiss, G. H. (1964). *Genetics*, **49**, 561–576.
- [25] Levins, R. (1969). *Bull. Entomol. Soc. Am.* **15**, 237–240.
- [26] Maruyama, T. (1969). *Genet. Res. Camb.* **15**, 221–225.
- [27] McKane, A. J. (2009). In: *Encyclopedia of Complexity and Systems Science*, (Meyers, R. A., ed) pp. 8766–8783. Springer New York.
- [28] McKane, A. J., Biancalani, T., & Rogers, T. (2014). *Bull. Math. Biol.*, DOI:10.1007/s11538-013-9827-4.
- [29] Milton, A. & Stegun, I. A., eds (1965). *Handbook of Mathematical Functions*. New York: Dover Publications.
- [30] Moran, P. A. P. (1957). *Math. Proc. Cam. Phil. Soc.* **54**, 60–71.
- [31] Moran, P. A. P. (1962). *The Statistical Processes of Evolutionary Theory*. Oxford: Clarendon Press.
- [32] Nagylaki, T. (1980). *J. Math. Biol.* **9**, 101–114.
- [33] Nagylaki, T. & Lou, Y. (2008). *Tutorials in Mathematical Biosciences IV*, volume 1922 of *Lecture Notes in Mathematics*. Berlin: Springer.
- [34] Nowak, M. A. (2006). *Evolutionary Dynamics: exploring the equations of life*. Cambridge, Massachusetts: Harvard University Press.
- [35] Risken, H. (1989). *The Fokker-Planck Equation - Methods of Solution and Applications*. Berlin: Springer, Second edition.
- [36] Rousset, F. (2004). *Genetic structure and selection in subdivided populations*. Oxford: Princeton University Press.
- [37] Sabin, L. (2009). *J. Math. Biol.* **59**, 659–696.
- [38] Serra, R., Andretta, M., Compiani, M., & Zanarini, G. (1986). *Introduction to the Physics of Complex Systems*. Oxford: Pergamon Press.
- [39] Tachida, H. & Iizuka, M. (1991). *Genet. Res. Camb.* **58**, 243–245.
- [40] van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier Science, Third edition.
- [41] Whitlock, M. C. & Barton, N. H. (1997). *Genetics*, **146**, 427–441.
- [42] Whitlock, M. C. & Gomulkiewicz, R. (2005). *Genetics*, **171**, 1407–1417.

- [43] Wiggins, S. (2003). *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. New York: Springer.
- [44] Wright, S. (1931). *Genetics*, **16**, 97–159.